INTRODUCING SYSTEM-BASED SPATIAL ELECTRICITY LOAD FORECASTING

MEHDI KARDEHI MOGHADDAM

THIS THESIS IS PRESENTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY AT MURDOCH UNIVERSITY

OCTOBER 2016

DECLARATION

I DECLARE THAT THIS THESIS IS MY OWN ACCOUNT OF RESEARCH AND CONTAINS AS ITS MAIN CONTENT WORK WHICH HAS NOT BEEN PREVIOUSLY SUBMITTED FOR A DEGREE AT ANY TERTIARY EDUCATION INSTITUTION

MEHDI KARDEHI MOGHADDAM

ABSTRACT

The main motivation of this research is to help reduce the Green House Gases (GHG) emissions of the electricity sector, and counteract the effects on nature and people. Traditional methods of power planning are not optimised to achieve this, and only consider Capital Expenditure (Capex) and Operational Expenditure (Opex) reduction as their main objectives. Minimising GHG emissions is now an additional objective of power planning. One way of achieving this is by optimising the distance of generators to the loads to reduce the transmission losses, and also by harnessing the available regional sources of renewable energies and increasing their integration in the network. Efficient load forecasting methods, capable of describing the regional behaviours of the electricity consumption are developed in this research, and can provide priceless input to electricity planners. Such forecasting methods, known as spatial forecasting, can be used to extract short-term and medium-term information of the electricity consumption of different regions. This work also provides tools for making decisions about the most accurate way of pre-processing consumption data and choosing the most efficient forecasting procedure.

Chapter 1 talks about emissions of GHGs and their adverse effect on the nature. It introduces electricity sector as one of the major contributors of human made GHG emissions. It then describes the components of electrical power network and the planning of it. Finally the chapter concludes that an efficient spatial load forecasting method is required to help with spatial planning of power networks. The spatial planning can include more regional components like proximity of generation components to consumers, or the levels of harnessed renewable energy in each area. In such an approach, GHG reduction can be also considered along with Capex and Opex minimisation to plan the future of power networks.

Chapter 2 provides definitions on power network components and the load forecasting methods. It starts with definition of power systems and explanation on how electrical energy is superior to all other forms of energy from end user point of view. Electricity generation systems and the sources of energy to produce electricity are described next. Typical generation unit sizes in MW, continuity of the supply, and also its predictability are summarised in a table at the end of this section. Thereafter, transmission lines and distribution systems are described, as other component of electrical power networks. Importance of having an accurate forecast of electricity demand and the common ways to do it are presented next. At the end of this chapter, the deficiencies of current forecasting methods are highlighted and one major goal is defined for this work. It is to overcome the deficiencies of individual forecasting methods by combining them and using them only where it performs efficient. It also mentions that the work is going to closely look at the behaviour of input data to the forecasting method to seek better methods for preparing them.

Chapter 3 describes South West Interconnected System (SWIS) as the case study for this work. The reasons for selecting SWIS as the case study are mentioned, followed by a quick history of it and how it has been expanded over the last hundred years. To be able to complete spatial forecasting, the area under study needs to be divided into regions. SWIS is then divided into eight regions for this purpose. A visual presentation of the eight regions on the map is presented at the end of this chapter for more clarity.

Chapter 4 performs a short forecasting method on one of the SWIS regions. The selected region is called Metro East. Metro East region is mainly composed of residential consumers. Unlike commercial and industrial consumers, the residential ones are not following a working schedule. That's why it makes them to behave differently and more randomly comparing to the other two. This means more complicated demand to forecast. This is the main reason that Metro East is selected to be studied on this chapter. One of the main components of this chapter is to introduce the methods that have been used for pre-processing of input data. The pre-processing stages include data resolution adjustment, replacement of missing data, removing outliers, clustering and signal reconstruction. A well pre-processed set of data is critical component of any forecasting strategy. The second component of chapter 4 is to generate one day ahead and seven day ahead forecasts of Metro East electricity consumption, using three different training methods. The forecasted results are comparable to other studies done on short term load forecasting. However the author questions the accuracy of classic approach of load forecasting. Classic approach is basically what have been done in the field of load forecasting for decades, which is very similar to the works done in chapter 4. In classic approach, a method gets tested on a case study with an acceptable level of accuracy. Then that method gets introduced as a very accurate tool to be applied on demand forecasting purposes. This work is showing that such accurate method cannot be accurate at all when being applied to other different case studies. Future chapters study this in further details, and come up with some guidelines on how to have accurate load forecast based on the nature of the case study in hand.

Chapter 5 applies the methods of load forecasting developed in chapter 4 onto eight different case studies. By doing this, it can be seen that there is no single method of forecasting that can be accurate for all case studies out there. Temperature sensitivity and distribution of the load data of all the regions is closely studied for fifteen years of data. A load type determination criterion is presented in Table 5. By using this table, and preparing Rayleigh, Generalised Pareto, and Generalised Extreme Value distributions of the load data under study, anyone will be able to say whether their load under study is mainly commercial, residential or industrial. The outdoor temperature is one of the main inputs of short term electricity forecasting. Same chapter shows that residential loads are having a greater temperature sensitivity comparing to the other two. The results of one day and seven day ahead forecasts of the eight regions are presented at the end of chapter 5, using two methods of neural networks and decision trees. The results suggest that the two methods need to be used alternatively based on the characteristics of the case study and ambient temperature to achieve the best result.

Chapter 6 explains the system based medium term load forecasting. The approach to medium term forecasting is completely different to the one developed for the short term one. Two main differences between Short-Term Load Forecasting (STLF) and Medium-Term Load Forecasting (MTLF) are the availability of weather data and the forecasting objectives. Because of the nature of the weather, temperature forecasts of a year ahead are completely impossible. Also in medium term load forecasting the focus of planners is mainly on peak load and energy

consumption forecasts. The forecasting method presented in this chapter is achieved by superimposing annual trend, annual seasonality and forecasted residuals by neural networks and decision trees. Similar to chapter 5, the forecasting strategy is applied to eight different case studies for comparison. It is concluded that based on the case under study, the accuracy of the methods changes. It also provides some advices on the best practices to perform medium load forecasting, considering the characteristics of the load. For instance, it conclude that for industrial regions regression trees performs better than neural network based methods. The same applies to CBD region where commercial load dominates. For some residential areas neural networks behave better. This is because of higher nonlinearity of residential load.

The major contributions of this work can be summarised as below:

- The topic of the study, i.e. spatial load forecasting and the potential of using it in efficient power planning, is relatively a new topic in the electricity market literature. Moreover, many of the known spatial load forecasting methods have not yet been widely used because of the size, variety, and availability of the data required. The methodology proposed in this study can successfully be applied to spatial forecasting.

- While conventional methods are useful for short-term predictions with acceptable accuracy, they fail when medium-to-long term load forecasting is dealt with. The methodology conceived and implemented in this thesis is significantly better than those known as state-of-the-art and can give very satisfactory results for medium-term predictions.

- The load analysis criterion, particularly using Q-Q (Quantile vs. Quantile) plots is a unique and original finding of this work. While Q-Q plots are largely used in traditional statistics to compare two samples of data, it has never been applied before for electricity load forecasting purposes. Based on its definition and use, an electricity planner can understand which part of the load is the dominating factor (i.e. whether it is residential, commercial or industrial). And then, based on this, he/she can decide how to go ahead with choosing the most effective forecasting method. Based on this, the thesis provides a very useful criterion for decision making in the energy market.

- One of the major findings of the thesis is that there is no one optimum way of forecasting electricity load in different scenarios. The results presented in the thesis have shown that a method that can accurately forecast the load on a system (3% error for a year ahead) can perform completely different in forecasting another system (observed errors of around 14%). This study demonstrates that a method which is claimed to have a given accuracy can be considerably inaccurate when applied on a different case study.

- Using an ambient temperature-based criterion (i.e. the average maximum temperature of the month) to choose the correct forecasting method is another major finding of the study. In fact, the author has demonstrated that for a temperature sensitive load, different forecasting methods should be used and then combined to get the most accurate result.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

ACKNOWLEDGEMENT

I would like to express the deepest appreciation to my supervisor Professor Parisa A. Bahri. She has been very supportive during the whole process of my PhD studies from early stages of preparing the research proposal all the way through to preparing the thesis. She has shown me how to be an independent researcher and how to face and tackle big problems all by myself. This will definitely be a priceless experience for me in my entire life.

Secondly I would like to thank my three years old son Mazdak. Daddy had to sacrifice a lot of his playing time with him to complete his PhD studies. Same goes to my wife Mahnoosh. Thanks for being an understanding partner.

A big thank you to Western Power for providing me with their resources and the opportunity to physically go to their office and gather the information I needed for my work. They also provided me with a comprehensive training on how to use their databases.

I would also like to thank Murdoch University for providing me with a generous scholarship to help me complete this work.

Finally I would like to thank my mother Masoumeh and my Dad Hassan for giving me an endless motivation during my entire life including my PhD studies.

Chapter

1 MOTIVATION

1.1 ABSTRACT

This chapter provides a motivation for this thesis. It includes a brief background on climate change, and introduces global warming as a probable source of future disasters. The chapter classifies anthropogenic greenhouse gases[1] (GHGs) [3] according to their sources of emission, and describes the electricity sector as the dominant source of GHG emissions globally. It concludes that anthropogenic GHG emissions of electricity sector should be limited and controlled. To achieve that, a reliable electricity planning strategy based on realistic load forecasts is required. The wider scope of the following chapters is given at the end of this section.

1.2 CLIMATE FACTORS

During the daytime about 1,370 Watts of solar radiation reaches each square meter of the atmosphere's outer surface. Only one quarter of this amount of energy enters the earth's atmosphere (see Figure 1), and the remaining is reflected back to space. A large amount of Earth's radiation is absorbed and reflected back to the earth from the atmosphere because of the existence of GHGs that cause a natural 'greenhouse effect', keeping Earth's surface relatively warm. The most significant GHGs are water vapour (H_2O) and carbon dioxide (CO_2), and the minor GHGs include methane (CH_4), nitrous oxides (NO_x), ozone (O_3), carbon monoxide (CO), and chlorofluorocarbons (CFCs). Human activities, such as, burning fossil fuels and deforestation, influence the natural greenhouse effect and are known to cause global warming[4]–[6].

1.3 CONTRIBUTION OF ANTHROPOGENIC EMISSIONS

Emissions of GHGs resulting from human activities are known as anthropogenic emissions. Radiative forcing is a parameter that shows the rate of energy change per unit area of globe measured at the top of the atmosphere [7]. Positive net radiative forcing shows increasing temperature of Earth-atmosphere system and negative net radiative forcing shows the opposite. Figure 2 shows principle components of climate change and their corresponding driving forces [6], [8].

[1]Greenhouse gases are atmospheric gases, both natural and anthropogenic, which absorb and emit radiationat specific wavelengths within the spectrum of infrared radiationemitted by the Earth's surface, the atmosphere, and clouds. Thisproperty causes the greenhouse effect. Water vapour (H_2O), carbondioxide (CO_2), nitrous oxide (N_2O), methane (CH_4), and ozone (O_3) are the primary greenhouse gases in the Earth's atmosphere. Moreover, there are a number of entirely human-made greenhouse gases in the atmosphere, such as, the halocarbons and otherchlorine- and bromine-containing substances, dealt with under the Montreal Protocol. Besides CO_2, N_2O, and CH_4, the KyotoProtocol deals with the greenhouse gases sulphur hexafluoride (SF6),hydro fluorocarbons (HFCs), and per fluorocarbons (PFCs)[133].

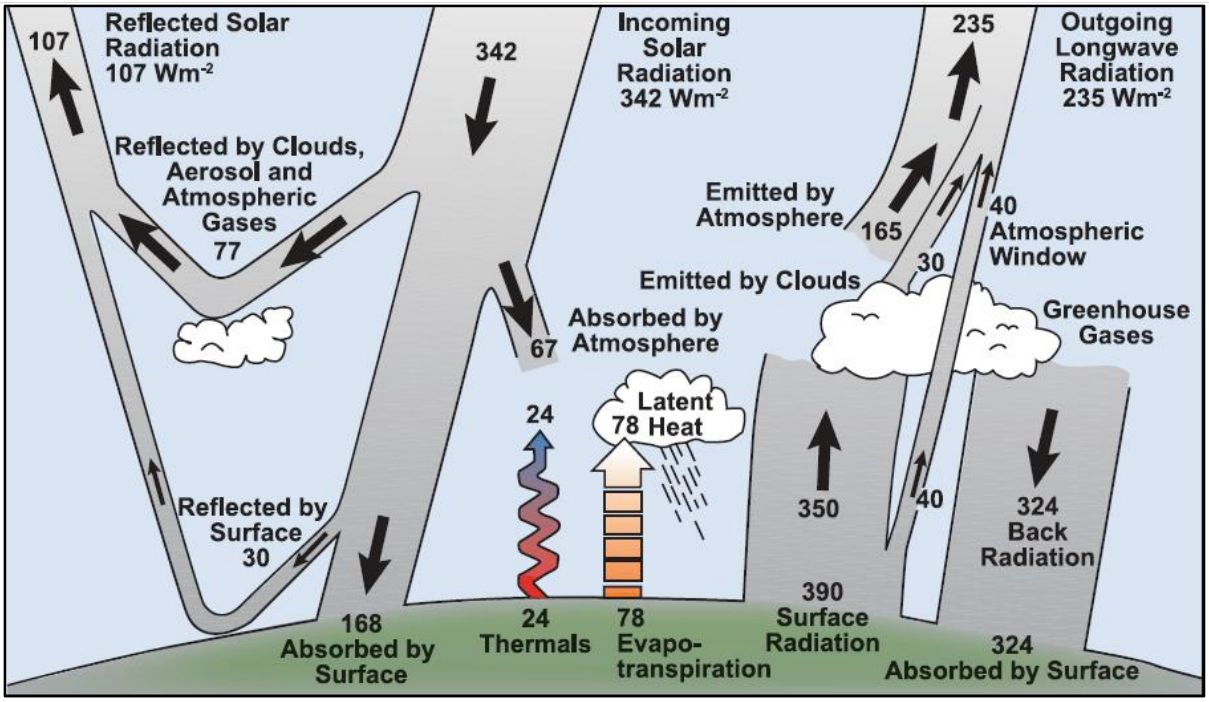


Figure 1: Earth's average energy balance(extracted from[9]).

Human activities are increasing GHG concentrations in the atmosphere[10], [11]:

- CO_2 from burning fossil fuels in the electricity, transport, industrial (etc.) sectors, and land use change (LUC) [12]. As illustrated in Figure 2, CO_2 emissions are the biggest contributor to the positive radiative forcing caused by human activities.
- AtmosphericCH_4 from agricultural related activities, livestock raising, losses from gas distribution systems, landfills, etc [13].
- NO_x as a fossil fuel burning by-product and from fertilizer use [14].
- Halocarbons from CFCs as refrigerants [15].
- Atmospheric O_3 from releasing CO, hydrocarbons, and NO_x which react to produce ozone [7].
- Atmospheric water vapour emissions, such as, evaporative cooling systems [16].

Figure 3 shows the anthropogenic emissions of different sectors. Electricity with 27% share plays as the biggest contributor of all the sectors in the level of global GHG emissions. It follows by other big contributors like land-use change, agriculture, energy sector rather than electricity, transportation and manufacturing and construction. The author's main motivation of this work is to help in reducing the GHG emissions of electricity sector by introducing more efficient tools of electricity forecasting and planning.

1.4 POWER SYSTEMS COMPONENTS

Power systems are composed of four main interconnected components [17]:

- Demand: Electricity demand is defined as the amount of electricity required to serve existing electrical loads in the network [18]. Regions of high populations or a high degree of industrial activity require more electricity than low populations or low levels of industrialisation. Electricity demand profiles also change with time [19].
- Generation: Electricity can be generated from different sources of energy and form the supply side in an electrical network. Examples are coal fired power plants, gas turbines, combined heat and power (CHP), piston-engine-based power plants, fuel cells, hydro power, tidal power, wind power, geothermal power, solar powers, wave power, biomass-based power, and nuclear power [20].
- Transmission: Generated electricity of power stations is normally in the range of 11 kV to 21 kV. To decrease the losses of lines, where electricity needs to travel a long distance, step up transformers are used to increase the voltage to 138 kV or 230 kV. Transmission lines are used to transmit electricity at this level of voltage towards consumers. This electricity needs to be stepped down to lower voltages of 11kV to 69kV when it gets close to the location of consumers [21], [22].
- Distribution: A distribution system provides the connection between transmission system and consumers [22].

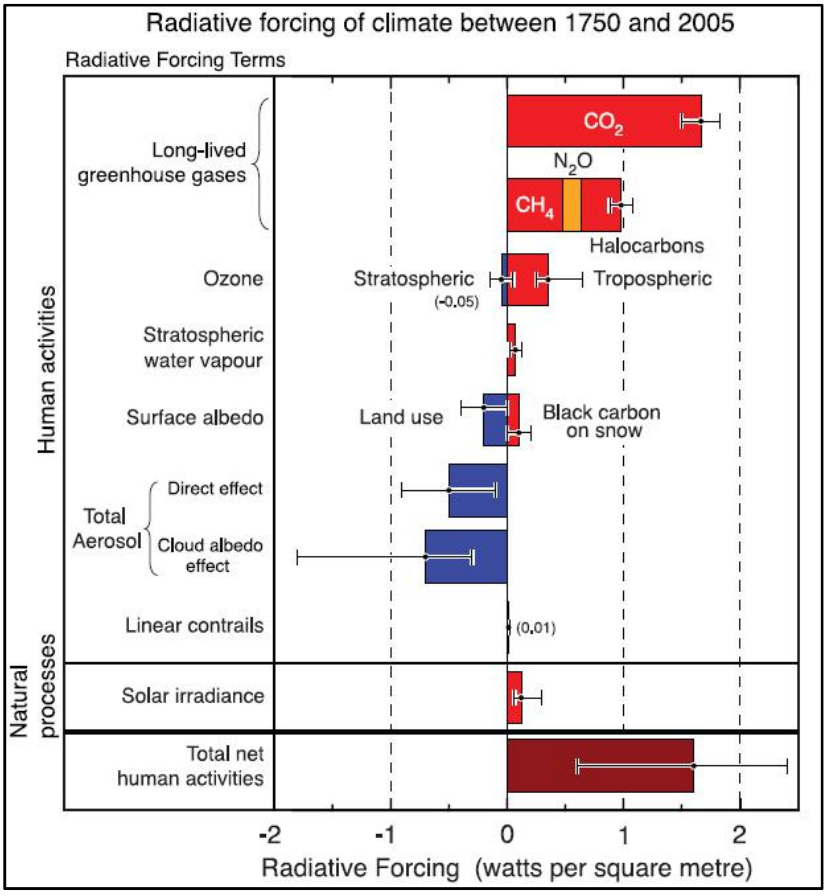


Figure 2: Principle components of radiative forcing of climate change(extracted from [6]).

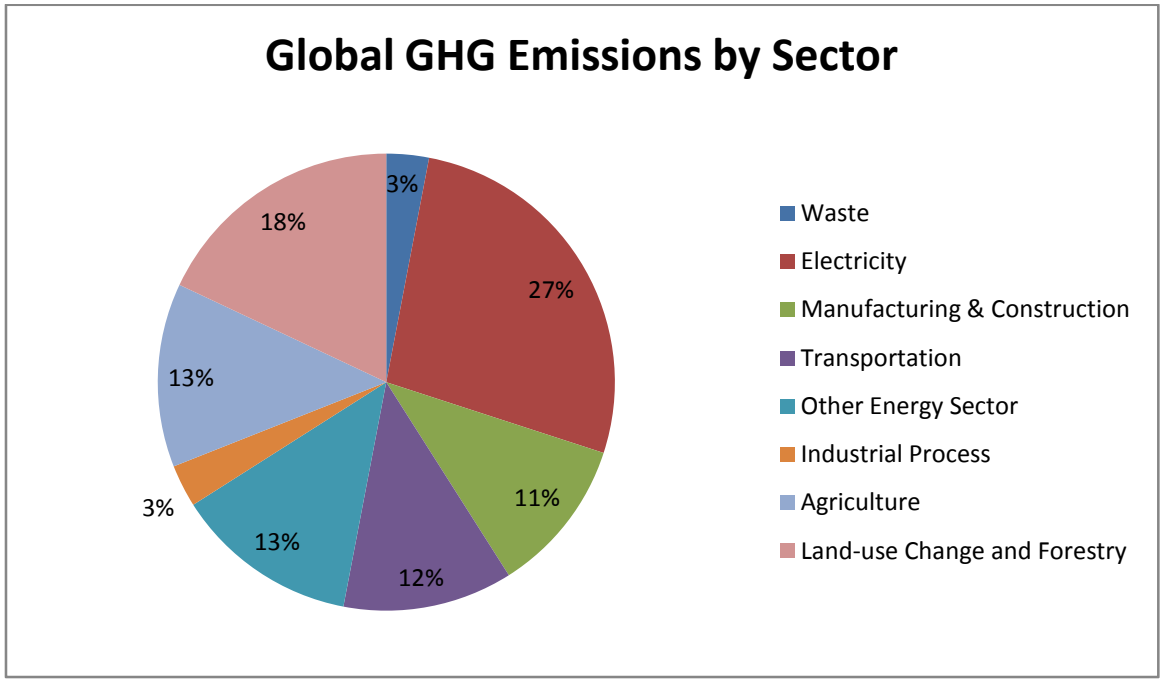


Figure 3: Global greenhouse gas emissions by sector(extracted from[23]).

1.5 POWER SYSTEM PLANNING

Power system planning encompasses generation planning, transmission planning, and distribution planning. The three planning procedures are very similar in steps and objectives but they consider different components of the electricity system [24]. The objective of power system planning is defined as minimisation of expansion cost conforming to financial, resource, political and environmental constraints.[2] Four different questions need to be answered in a power system planning procedure:

- WHAT capacity is to be installed to cover the demand in a reliable manner? [25], [26]
- HOW to select the best combination among the available and future technologies? [27]
- WHERE to place new facilities? (Facilities mainly consist of generation units in power generation planning, transmission lines and substations in power transmission planning, and distribution lines and substations in power distribution planning.) [28], [29]
- WHEN is the desired time to commission new facilities? [27]

[2] Although cost constraints are the main objective of this definition, other variables are assumed as constrained. A planner may add other main objectives to the problem, such as, minimising the GHG emissions, or minimising the space between generation facilities and the available resources.

Planners have proposed different answers to the above questions. Historically, the least attention was paid to the 'WHERE' question, and forms the basis of spatial load forecasting described in chapters 5, 6, and 7.

The main steps in power system planning are as follows:

1) Taking into consideration the most reliable forecast of future demand data for the specific planning horizon. Short-term horizons are used for day-ahead scheduling and real-time grid management. Medium-term horizons are used for preparing maintenance schedules for generation and transmission equipment, planning the most effective use of limitedly available resources, and the pooling of resources with neighbouring utilities. Long-termhorizons are used for infrastructure additions [24].
2) Studying the future available resources and technologies for producing and transmitting energy and their related economics [30]–[32].
3) Investigating economics of present and future facilities including capital cost, operating cost, fuel cost and maintenance cost [33].
4) Determining the internal rate of return, level of reliability, investment policies [33].
5) Preparing the expansion scenarios considering constraints [27].
6) Qualifying the output scenarios to check the solution viability[27].

Forecasting of future parameters for power system planning is a challenging task[34]. Future demand and fuel costs are the main sources of uncertainty[35], and must be determined as a prerequisite for power expansion planning. Accurate load forecasting can improve electricity system planning for spinning reserve[3], generation planning, scheduled utility maintenance, generation cost, and system reliability[36].

1.6 LOAD FORECASTING

Load predictions are required for short time intervals for operational purposes, or long intervals over decades for power generation planning procedures [36]. Factors that influence future loads are GDP,[4] population, ambient temperature, and humidity. The level of correlation among these factors and the electrical load varies for different electricity systems[37].

More detailed information on electricity demand forecasting will be given in future chapters.

1.7 FOCUS AND OUTLINE OF THIS PROJECT

Figure 3 presents the electricity sector as the largest contributor (27%) to global anthropogenic GHG emissions. The main motivation of this work is to help reduce the GHG emissions of the

[3]The extra generating capacity available by increasing the power output of generators that are already in use in a power system is known as spinning reserve.
[4]Gross domestic product is a factor of country's overall economical output. GDP is closely related with the standard of living. It is the market value of all final goods and services made within a country in one year[134].

electricity sector. Traditional methods of power planning consider Capex[5] and Opex[6] reduction as their main objectives [38]. Minimising GHG emissions is now an additional objective of power planning [1], [2]. One way of achieving this is by optimising the distance of generators to the loads to reduce the transmission losses and also by harnessing the available regional sources of renewable energies and increasing their integration into the network [39]. The development of an efficient load forecasting method capable of describing the regional behaviours of the electricity consumption is a useful planning tool. Such forecasting method is known as spatial forecasting [40], which is discussed in Chapter2. Spatial power planning is mainly applied in distribution networks and for short-term planning purposes[40]. This work is going to look at the spatial forecasting of the electricity consumption at a larger scale. The output of the work would be useful for developing more efficient transmission and generation power plans. Below is a description of the content of upcoming chapters.

Chapter2 presents a comprehensive literature review of power system components and load forecasting.

Chapter 3 presents south western Australia electricity network known as SWIS[7] as our case study.

Chapter 4 discusses short-term load forecasting based on classic approaches.

Chapter 5 introduces a new means of short-term load forecasting using an innovative system based approach.

Chapter 6 extends the ideas developed during chapter 5 and modify them to suite spatial medium-term load forecasting.

Chapter 7concludes the research with some potential applications and recommendations.

Appendix Ais the bibliography.

Appendix B presents some information on Western Australia public holidays that is used to consider the effect of public holidays in electricity load forecasts.

Appendix C are Australian distribution figures used to define sub-regions.

Appendix D presents selected MATLAB codes used in this thesis.

Appendix E is the nomenclature of this work.

Appendix F is the glossary of the-terms used in this work.

[5] Capex or capital expenditure are the funds used by companies to purchase physical assets such as buildings, power plants and equipment.

[6]Opex or operational expenditure is the money that a company spends during its business as usual to cover its day to day costs. Examples are fuel costs, equipment rentals and salaries paid to staff members.
[7]SWIS is the-term used to describe southwest Australian interconnected electricity network. Refer to 3.2 for more details about SWIS.

Chapter

2 LITERATUREREVIEW

2.1 ABSTRACT

This chapter explains the concepts of power systems and load forecasting in more detail based on the available knowledge and requirements of this work.[8] A brief description of the case study in this work and the reasons that it has been chosen are provided, followed by an explanation of spatial load forecasting.

2.2 POWER SYSTEMS

An electric power system is a network of electric components that supply, transmit, and consume electricity. Generation, transmission, distribution, and demand are the main components of an electric power system [17].

The world's first power system was built in 1881 in England. The generation stations were two power wheels and the application was to turn on a number of lights. The power source was intermittent and the output was alternating current (AC) [41]. Due to advancements in power systems technologies and applications since 1881, it has become an essential part of life in the modern world.

Electrical energy is superior to all other forms of energy for the end user due to the following advantages [42]:

- Convenient form: It can be easily converted to other forms of energy.
- Easy control: Electric energy can be controlled by very basic circuit components.
- Great flexibility: With the help of conductors it can be easily transported.
- Cleanliness at the point of use: Given that consumers of electrical energy are usually situated far away from the electricity production sites,[9] they experience no smoke, fumes, or poisonous gases.

Figure 4 shows a very simplified diagram of a power system. Generation, transmission, distribution, and customers or loads are identified in the diagram. The working voltages of each section are also shown [43]. More details of power systems components are discussed in future sections.

2.2.1 GENERATION SYSTEMS

The generation units generate electricity from different sources of energy (such as, coal, gas, sun, wind, and distillates), and form the supply side in an electrical network. Power generation units can be a huge power plant or a small solar panel installed in a house.

[8] As this work is focused on electricity demand forecasting, the literature review on other subjects gives an introduction of some power systems concepts to provide additional context for the reader.

[9] Exceptions can be places with small generators like diesel generators or factories that generate their own electricity.

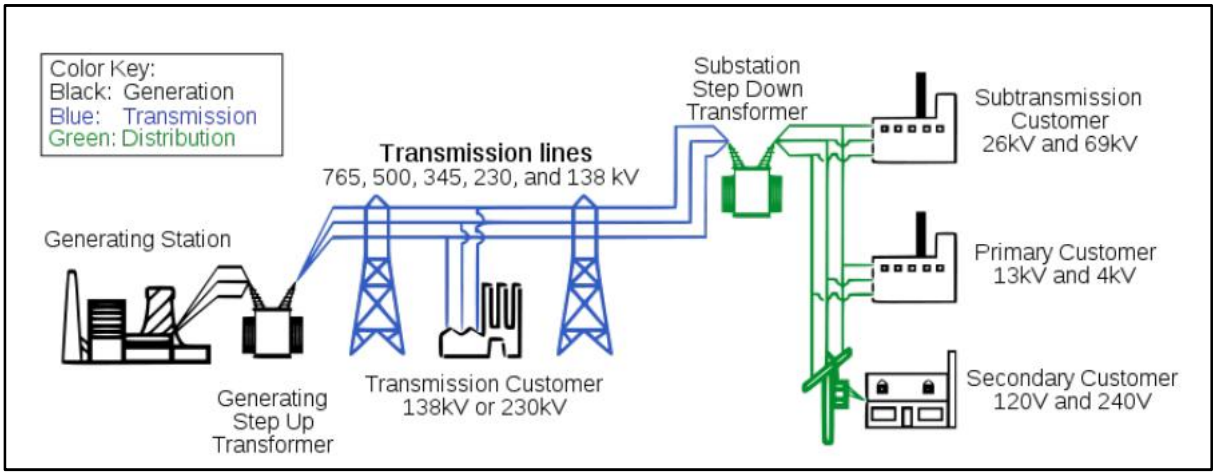


Figure 4: A simplified diagram of a power system(extracted from [43]).

Electrical energy can be generated from either renewable or conventional sources of energy [44]. A renewable resource has the ability to reproduce through biological or natural processes and renew with the passage of time. These sources can be harnessed to generate sustainable electricity [45]. In-terms of availability and abundance, solar energy is the most accessible renewable energy. The amount of solar energy resource available on the surface of the Earth is around 8000 times the world's average energy requirements. This energy can be accessed directly using solar thermal or photovoltaic technologies, or indirectly from wind, wave, hydro, and biofuels [46]. Two other sustainable sources of energy are the gravitational forces of the moon and sun, known as tidal energy, and the Earth's core thermal energy, known as geothermal energy [47]. Figure 5 represents a summary of renewable sources of energy and their related renewable energy (RE) technology.

Summary information on RE generation is provided below [47]–[62]:[10]

- Wind power: Wind is characterised as a free, intermittent, clean, inexhaustible, and nondispatchable[11] source of energy. Wind is generated from the uneven heating of the earth's surface and atmosphere. Thus, it is known as an indirect form of solar energy. Although wind patterns on the Earth's surface are derived mainly from differential surface temperature and atmospheric pressure, the Earth's movement also affects wind direction around the globe, known as Coriolis. The pattern of the terrain is another factor that influences wind intensity and direction. The terrain pattern is characterised by items like surrounding trees, mountains, valleys, and buildings. Wind energy can be harnessed using wind turbines. The output power of a wind turbine is proportional to

[10] An extensive review of renewable energies is beyond the scope of this thesis. Please see the sources for further information on renewable generation.
[11] Dispatching is the planned allocation of a generating plant to meet future loads on the system. Most of the renewable sources of electricity generation are considered nondispatchable because their generation is directly affected by natural variations in the reliability and strength of the source.

the cube of wind speed. Wind variability is the main challenge in designing wind farms and the integration of this source of energy into the electricity grid.

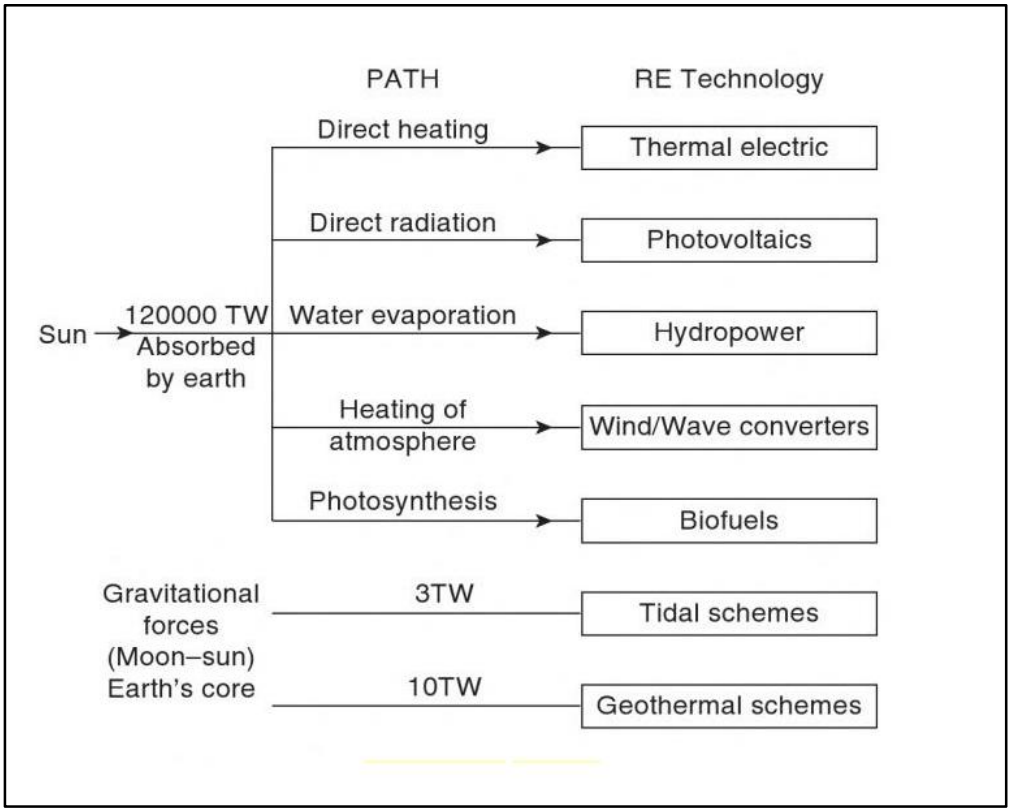


Figure 5: Renewable energy flow paths(extracted from [47]).

- Photovoltaic and solar thermal electricity: Solar radiation is electromagnetic radiation generated by the sun in wavelength ranges from X-rays to radio waves. The-term 'solar radiation' in renewable energy applications is the optical range of solar radiation belonging to the spectral range of 300-400 nm. Solar radiation received by a surface is made up of different components (see Figure 6). The components are direct beam radiation, reflected radiation from other objects, and diffuse sky (refer to section 1.2for a detailed discussion on solar energy components). Solar radiation patterns are more regular compared to wind speed patterns. There are two main RE technologies to convert sun radiation into electricity. Photovoltaic (PV), the most direct way of converting sunlight to electricity, is based on the electric voltage generated between two attached electrodes in a solid or liquid system upon receiving optical radiation. Although the efficiencies of commercial PV cells are in the range of 12-18%, much better efficiencies (over 30%) have been achieved on experimental cells. In contrast, solar thermal systems convert solar radiation into heat. The heat is extracted from the short-wave sun radiation absorbed using a black-coated surface. This thermal energy can be

used to form steam and drive a turbo-generator to generate electricity.[12] Both of the mentioned methods are relatively expensive, in-terms of capital costs, compared to other available methods of electricity generation.[13]

- Hydropower: The first electricity generation plant was built in 1881 using two water wheels with hydropower turbines using the same concept today to generate electricity. Hydropower plants generate about 24% of the world's electricity supply. The potential energy stored from water elevation in rivers, waterfalls, and dams, known as head, is an indirect form of solar energy. To clarify, solar energy causes the water on Earth to evaporate and form clouds, thus converting solar energy into mechanical potential energy. Latent heat is released when precipitation occurs, although some of the potential energy remains with water when it falls on high elevations. Rain drops that land on high elevation surfaces, like mountains, make rivers and the called head difference to power hydroelectric power plants.
- Wave power: Wave energy can be defined as a concentrated form of solar energy. As previously discussed, wind is generated from the sun. The movement of wind over the ocean surface makes ripples, which then grow into swells, forming waves. Deep water waves can travel kilometres without any loss of energy. The concentration of this source of energy is greater than solar and wind energy. The power content of a wave is proportional to its period and square of its height. Although several technological solutions have been developed to harness wave power and convert it to electricity, these technologies are still in the earlier developmental stages when compared to wind and solar technologies.
- Biomass: Biomass is very different from other renewable sources of energy. It is a fuel obtained from wood, straw, twigs, dung, agricultural residues, etc., and is used to generate electricity. The world's biomass resources can contribute over twice of the world's energy demand. Due to the low energy density and high transportation costs, biomass should be consumed locally to be beneficial. Three main technologies are available to generate electricity from biomass, namely direct combustion, gasification, and pyrolysis.
- Tidal Power: Tides are the vertical rise and fall of the oceans' water. Rotation of the Earth around its axis generates large centrifugal forces causing the Earth's diameter at the equator to be 21 km larger than through the poles. For the same reason, the depth of the seas at the equator are also greater than at the poles. Because the speed of rotation around the Earth's axis is constant, sea depth differences do not generate tides. Instead, tides are formed by the rotation of the Earth within the gravitational fields of other planets and the sun, specifically, the gravitational fields of the moon and sun. Tidal barrage can be used to harness the energy from the rise and fall of the tides. However, long payback periods and large capital expense requirements discourage widespread investment in this technology. Furthermore, although tidal power barrage is considered

[12] The main current application of solar thermal systems is to produce domestic hot water, not electricity generation.
[13] Renewable energy technologies are becoming cheaper over time because of the advancement of technologies and mass production.

non-polluting, the resulting changes in tidal patterns have significant environmental effects.

- Geothermal power: Geothermal energy is the thermal energy from underground sources of hot water and steam, and can be used to generate electricity. In some cases, they are as visible as hot water or steam springs. The basics of geothermal power plants are very similar to other power plants except that they do not burn fuels to generate electricity. In a typical geothermal plant cold water is pumped underground to exchange heat with geothermal reservoirs and to make steam required to propel turbines.

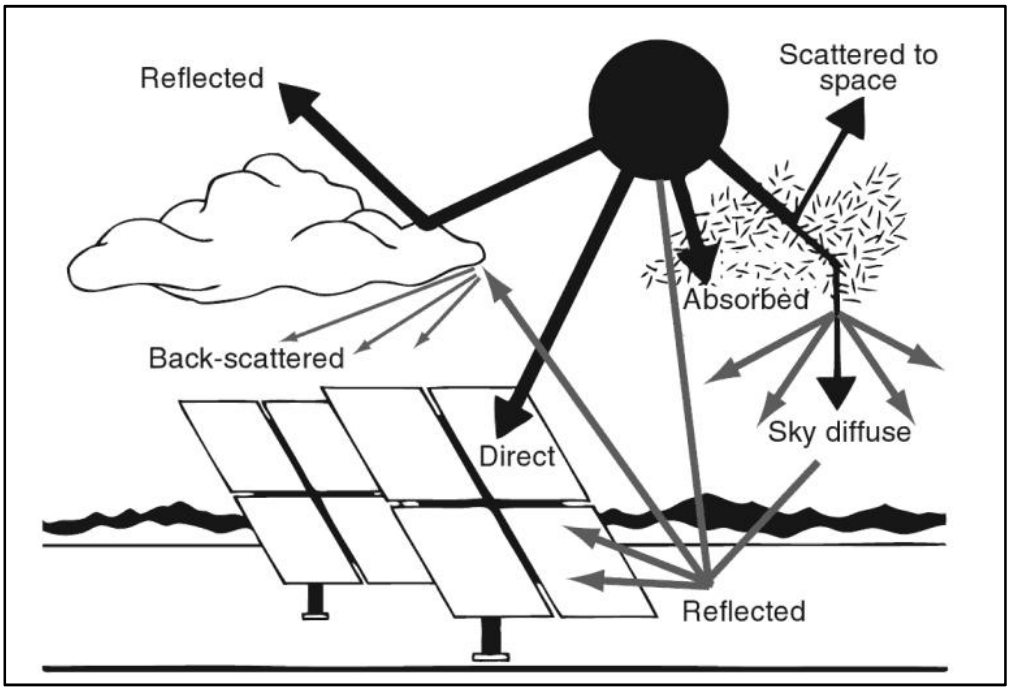


Figure 6: Solar radiation components(extracted from [55]).

In contrast to renewable energy resources, non-renewable sources of energy consist of a natural source, once depleted, cannot be reproduced for future use. Fossil fuels and nuclear power are the main sources of non-renewable electricity generation, as described below [63]–[66]:

- The most common way of producing electricity is by burning fossil fuels (coal, oil, or natural gas) in a boiler to produce steam to turn a turbine mechanically connected to an electrical generator. There are three main reasons to seek alternatives for the use of fossil fuels in energy generation. Firstly, as a non-renewable resource, fossil fuels are finite, and therefore will one day be unavailable as a form of energy generation. The second issue with fossil fuels is the geographic location of current reservoirs, which is controlled exclusively by a limited number of governments in-terms of both price and availability. Finally, contaminants are produced through the burning of fossil fuels, which severely affect the environment.

- In 1905, Albert Einstein discovered that when atoms break apart the mass of atoms can be converted to a massive amount of energy. Scientists later found that 0.45 kg of radioactive uranium can produce energy equal to burning 1.4 million kg of coal. The process of breaking atoms apart is called fission. The first attempts of harnessing nuclear power were to make weapons with the first nuclear bomb tested by the United States of America in New Mexico during the 1940s. By the 1950s, efforts were focused on controlling nuclear power to generate electricity, leading to the construction of the world's first commercial nuclear power plant in England in 1956. A nuclear reactor, required to harness nuclear power, is equivalent to a boiler in a fossil fuel generation plant. The main components of a nuclear power reactor are: a nuclear fuel core; a system to control the fission rate, and; a cooling system to keep the fuel from overheating. Although the harnessing of nuclear power to generate electricity is almost CO_2 free, it indirectly impacts the environment through water pollution, radioactive dust, radioactive gas, and other radioactive waste, such as, chlorofluorocarbons.

The key characteristics of different sources of electricity generation, presented in Table 1, shows the typical size of a single generator unit, output variation in time, the output predictability, and whether or not the generator is dispatchable.

Table 1: Generator characteristics by energy source(extracted from [47]).

Energy source	Typical unit size	Variable	Predictable	Dispatchable
Coal	500 MW	No	Yes	Yes
Nuclear	500 MW	No	Yes	Yes
Gas CCGT[14]	Up to 500 MW	No	Yes	Yes
Gas open cycle	100 MW	No	Yes	Yes
Hydro with reservoir	Up to 500 MW	No	Yes	Yes
Pumped storage hydro	Up to 500 MW	Yes	Yes	Yes
CHP[15]	Up to 100 MW	Usually	Usually	No, because it is heat led
Energy crops and municipal solid waste	Up to 40 MW (at present in UK), in future larger units	No	Yes	Yes
Wind	Up to 5 MW	Yes	Not accurately	No
Landfill gas	1 MW	No	Yes	Yes
Run-of-river hydro	100 kW	Yes	Not accurately	No
Photovoltaic cell	1 kW domestic, up to 100 kW commercial	Yes	Not accurately	No
Wave	No commercial examples yet	Yes	Not accurately over long-term	No
Tidal	No recent	Yes	Yes	No

[14] Combined cycle generation technology.
[15] Combined heat and power plant, also known as cogeneration.

2.2.2 *TRANSMISSION SYSTEMS*

The purpose of the electricity transmission is to deliver the electricity generated by power plants to substations located near to the loads. Although different methods of power transmission have been developed and tested, the most common way of transmitting electrical power is the AC system. High voltage direct current (HVDC) technology is mainly used for very long distances (greater than 600 km). The problem with AC transmission on very long cables is increased reactive power. Other applications of HVDC systems are submarine cabling of longer than 50 km and the connecting of two non-synchronised AC power networks [67]. Single phase AC systems are also used for electrifying railways [43]. Different countries use different frequencies, for example, 50 Hz in Australia, Europe, and parts of Asia, and 60 Hz in the USA. The most ubiquitous and economic means of power transmission are overhead power lines. Thus, the determination of line voltage drop, line losses, and the efficiency of transmission are important considerations in design and operation of transmission systems [48]. Overhead transmission lines are classified into three groups, according to line analysis requirements [42]:

- Short transmission lines are less than 50 km long with a line voltage of less than 20 kV. Due to the short length and relatively low line voltage, the capacitance effects of the line can be neglected and power engineers only consider line resistance and inductions for calculations.
- Medium transmission lines are between 50 km and 150 km with a line voltage ranging from 20 kV to 100 kV. In this case, the line capacitance is divided and lumped as capacitors shunted to the line at one or several points.
- Long transmission lines are longer than 150 km with a voltage higher than 100 kV. The line characteristics (including capacitance) are assumed to be uniformly distributed along the length of the line.

To decrease the power losses of transmission lines, the output voltage of power plants is usually increased[16] by a step-up transformer and transmitted to distant places using overhead lines. At closer proximity to the load locations for distribution, the voltage is decreased using step-down transformers. Transmission line voltages are normally higher than 110 kV [68]. Lower voltages like 66 kV and 33 kV are classified as sub-transmission voltages. Voltages lower than 33 kV are typically used for distribution purposes [43].

2.2.3 *DISTRIBUTION SYSTEMS*

The main differences between distribution and transmission systems are proximity to load centres and voltage levels. There is no universal distinction between transmission and distribution lines [69]. Generally speaking, transmitting the electricity from generation substations to regional depots can be considered transmission, transmission from the depot to central area warehouses is sub-transmission, the warehouses to local wholesale vendors forms the primary distribution, and finally, from vendors to local consumers is secondary distribution [70] (Refer to Figure 1 and Figure 4 for further clarification). The problem with this figure is the typical voltage levels. The definition of high voltage and low voltage lines changes over time,

[16] Based on electrical circuit rules, the power loss of a conductor is proportional to the square of current flowing through it. Because the line resistance is constant, a step-up transformer can reduce the current by increasing the secondary voltage and help to decrease the power losses of transmission lines.

that is, lines that are considered high voltage today, may be referred to as low voltage in 20 years time. Thus, it is preferable to distinguish between three main distribution system components (see Figure 8) from generation to consumers for use [42]:

- Feeders are conductors that connect sub-stations to the area where power is to be distributed. The main concern in the design of the feeders is carrying current capacity, which is usually constant in each feeder because no tappings[17] are taken from it. SA and SB are feeders in Figure 8.
- Distributors are conductors where tappings are taken to supply consumers. Obviously the current is not constant in this case and the main concern in distributor design is the voltage drop across its length. AB, BC, CD, and DA present distributors in the following graph.
- Service mains are cables that connect distributors to consumers' premises up to the metering point. Service mains are shown with small arrows onFigure 8.

2.2.4 DEMAND

The end points of every power network are consumers, who differ in patterns of demand from available electricity based on their different needs. Three types of consumers are examined within this thesis:

- Residential consumers: Demand patterns of residential consumers are the hardest to predict as each household uses electricity in a different way according to their needs and preferences [71]. In areas where electricity is being used for cooling and heating purposes, however, there will be a strong correlation between residential demand and ambient temperature [72].
- Commercial consumers: Commercial demand can also correlate with ambient temperature. However, after deducting the temperature dependant component of commercial demand, demand, in general, is easier to predict and remain fairly constant during working hours [73].
- Industrial consumers: Unlike two previous categories, industrial demand does not correlate strongly with ambient temperature. However, it is not easily predictable due to fluctuations in work load and overhaul periods [74].

2.3 LOAD FORECASTING

One of the critical tools of planning is to foresee the future. Forecasting of future parameters affects business activities such as: following up technological evolutions, revenue estimation, maintenance planning and replacement of major plant and equipment. To predict the electricity consumption of an electricity network in a specific point of time in future is called electricity load forecasting. This section reviews the methodologies and practices used for electricity load forecasting.

[17]A tap is a connection from the transformer winding that allows a specific number of turns to be selected.

In operating power systems, the main application of electricity forecast is to assure that enough supply is available to match the load. The range of forecasts can be from a couple of hours, known as short-term forecast, to long-term forecasts, which can be up to 30 years. Forecasting horizons from seven days to two years are considered medium-term forecasts [75].[18] This research is concerned with short-term and medium-term forecasts (long-term forecasts are beyond the scope of this thesis).

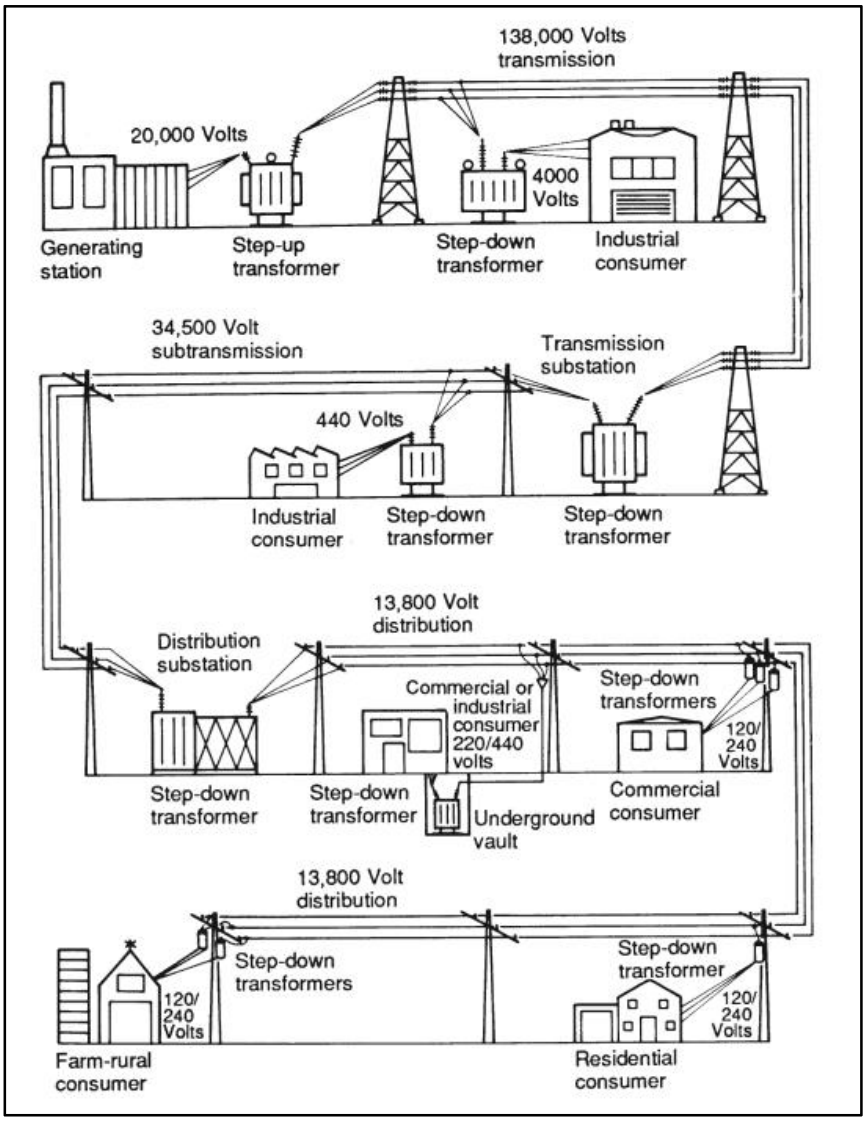


Figure 7: Typical operating voltages from generator to consumer(extracted from [70]).

[18] These definitions can be changed according to the planning requirements. The time horizons mentioned here are for the purpose of this thesis.

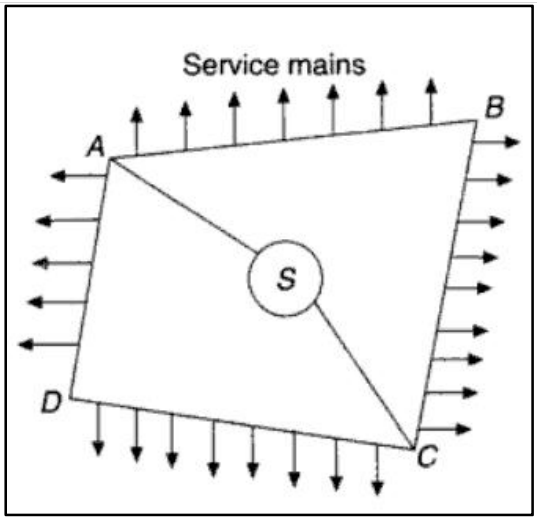


Figure 8: Distribution system components(extracted from [76]).

Long-term forecasts are mainly needed for capacity expansion plans and long-term investments. Medium-term forecasts are used for maintenance scheduling while short-term and very short-term forecasts are useful for operation planning, economic dispatching, load exchange and contracting with neighbouring networks [77].

An accurate load forecast is a vital part of each power system. Because energy storage systems at large scale are not yet very well applied on power systems, the most accurate forecast is required to help with short term investment decisions [75].

Electricity load forecasting approaches can be categorized into different groups based on factors such as: application type, the area of application (generation, transmission or distribution planning), the time frame (short-term, medium-term, long-term), and the tools to be used to complete the forecast [75].

There are two main methods of forecasting electricity load. One is statistical-based models and the other one is artificial intelligence-based models[78]. There is no specific preference on which methodology to use and it all depends on the application. Below is a brief explanation of these methods. More detailed comparison of these methods with examples will be discussed in future chapters.

2.3.1 Statistical-Based Models

Statistical-based approaches are widely used in different forecasting applications. They work very well under normal conditions but as it will be shown in future chapters they are not capable of capturing complex nonlinearities[79]. This usually happens with sudden changes in input data like weather parameters[80]. Two of the mostly used statistical models are regression models and time series.

2.3.1.1 Regression Methods

Regression is the study of relations between variables and is one of the most widely used methods of load forecasting. After working out the relation among variables the resulting

formula can be used to predict future variables. Input values can either be estimated or assumed. In load forecasting applications the inputs can be weather data, day of the week or holidays [80].

The regression equation can be either linear on nonlinear. By initial investigation of the data, the forecaster can decide which equation to go for. It can be linear, polynomial, exponential, trigonometric or other functions. The coefficients of the equations need to be solved to minimize the sum of squared errors. The model needs to be kept as simple as possible to make it easy to solve for the coefficients of the equations. Refer to [81] for traditional methods of linear regression, multiple regression, and exponential methods that have been applied for a day ahead forecasting.

2.3.1.2 Time Series Methods

Time series methods work well with the data sets having internal structure. This can be autocorrelation, seasonal behaviours or trends. Electricity load data do have these characteristics and therefore time series methods is one of the favourite methods of electricity load forecasting [80][82].

A time series model can be broken into below components. These components can be added for forecasting purposes[19].

- Trend component. This can be any trend (line, polynomial, exponential, etc) and represents the long-term pattern of the series.
- Seasonality: This component is when the data repeats itself in a given period of time. In load forecasting applications several seasonal behaviours are present (daily, weekly).
- Residuals: Residuals are the remainders of above components which present randomness in the system.

Methods named ARMA, ARIMA, ARMAX and ARIMAX are all different variations of time series. The following can be used to translate the meaning of each method:

- AR: Autoregressive
- MA: Moving average
- IMA: Integrated moving average
- X: With exogenous variable

Application of time series is presented in future chapters. For more examples the reader can refer to [83] which presents the seasonal auto regressive moving average (ARMA) and double seasonal Holt-Winter exponential smoothing.

2.3.2 Artificial Intelligence-Based Models

Historically many artificial intelligence-based methods were proposed to help with load forecasting applications but, because of the shortage of computational resources they were not easy to apply. Lately after introducing advancements in computer technologies, tools became

[19]Applications of such combinations are provided in future chapters.

capable of handling complex computations required to do artificial intelligence-based forecasts [84].

Below is a summary of the artificial intelligence-based methods mainly used for electricity forecasting purposes.

2.3.2.1 Neural Networks

Artificial neural networks have been used widely for load forecasting [85]. Artificial neural networks are generally used for non linear curve fittings. Neural networks can have different architectures. Examples are Hopfield, back-propagation and Boltzman machine. The most popular one in electricity load forecasting is back-propagation which is abbreviation for "backward propagation of errors". Back-propagation networks utilise constantly valued functions and supervised learning. After feeding a set of input and output data to the network it starts to optimize its weights. When the network is trained and the weights are set another set of input and output data can be used for model validation. When the validation is satisfactory the model can be used to generate outputs for the new sets of inputs[86]. In our case the inputs will be weather data, day of the week data, month of the year data, year and public holidays and the output will be the electricity demand. More details and applications of artificial neural networks are presented in future chapters.

2.3.2.2 Fuzzy Logic Systems

Fuzzy logic was introduced in 1965 by Lotfi A. Zadeh. Unlike classical logic, fuzzy logic does not assign a specific numeric value to each parameter. Qualitative rules are used instead to related inputs to outputs[87]. To simplify, for example, there is not an exact number to define whether something is cold or hot. Fuzzy logic can define a rule for such variable which is basically a curve to qualify this characteristic. To generate outputs a reverse procedure is used which is called difuzzification [88]. Fuzzy logic has been applied for daily peak forecasting in [89].

2.3.2.3 Support Vector Machines

Support vector machines SVMs are supervised learning methods to recognize patterns in data sets and are usually used for classification and regression analysis. Although they are mainly used for linear classifications they can be modified using what is called kernel trick to perform nonlinear classifications. SVMs have been used in forecasting electricity demands and attractive results are found compared to statistical methods [90]. Support vector regression based on statistical learning theory has been used for load forecasting in [91].

2.4 Conclusion

Many different methods have been used for load forecasting. Each of these forecasting methods has its own capabilities. Statistical based-models works well under normal operations but they are not very reliable in dealing with sudden changes in input variables like weather data. As stated in 2.3.2, artificial intelligence based models works well with system nonlinearities. They are very good for short-term load forecasting applications where randomness is small. In longer-term load forecasts where degree of uncertainty increases, the artificial intelligence-based methods become less efficient and will have large forecasting errors.

The author is after innovative methods to resolve the above deficiencies of different methods. This goal is going to be achieved in future chapters by combination of both statistical and artificial intelligence-based methods, and by intensive pre-processing of data and classifying the area into sub-regions to achieve a better classification of the data sets. Three methods of regression trees, neural network, and auto regression with exogenous variables (ARX) are applied in this work. These methods are chosen to further investigate the behaviour of each method in load forecasting. Artificial neural networks is chosen as representative for artificial intelligence-based methods, ARX is chosen as a representative of statistical-based methods. Regression trees as a regression method with if-then tree architecture represents a method that is both statistical-based and artificial intelligence-based.

The preparation of input data and selection of an appropriate optimization algorithm plays a critical role in accurate load forecasts, which is explored in more detail in Chapter 5.

Chapter

3 CASE STUDY

3.1 ABSTRACT

This chapter introduces SWIS as the case study of this work. It also represents the concept of meshing and the items to consider when preparing a mesh for electricity planning purposes. The visual presentation of the SWIS mesh is shown at the end of the chapter.

3.2 SOUTH WEST INTERCONNECTED SYSTEM (SWIS)

The South West Interconnected System (SWIS) was selected as the case study for this research. The SWIS is located in the South West of Western Australia and has some unique characteristics that make it an ideal case study for load forecasting studies, for example:

- The SWIS is an isolated network, thereby eliminating the effect of multiple networks on each other, making the regional study of load consumptions relatively straight forward [92].
- There are various types of regions with different load patterns present in the SWIS. These unique characteristics will be used in future in the thesis to test the developed platform on different sub-systems and test the effect of input behaviours on the results.
- The SWIS is highly temperature sensitive with electricity used to cool or warm residents. Thus, the load fluctuates dramatically by ambient temperature changes. This is important for the sub-systems with a higher residential component. In general, the greater the temperature sensitivity of a region, the greater the nonlinearities within the system, making the system more complicated to forecast.
- The SWIS network has experienced constant growth in the last 15 years, like most other electricity networks around the globe.

3.3 BRIEF HISTORY OF THE SWIS

The SWIS has developed to provide power to Western Australia. The network was originally expanded to satisfy the requirements of population growth and later on developed to support industrial activities associated with mining and export of raw materials.

The beginning of the SWIS was in 1913 after the merging of several independent power operations in Perth Area. By 1916 the first power station was built in Western Australia at East Perth with the total capacity of 12 MW. The power was provided to Perth using 6.3 kV lines. The total demand was below 4 MW at that time. The SWIS has constantly grown since then and the current SWIS is capable of managing the system peak load of 4028 MW[93].

3.4 Dividing the SWIS into Regions

To forecast the electricity consumption within a specified location, where the consumers are physically close to each other, is known as spatial load forecasting. There are few sources in the literature on spatial load forecasting. The idea of spatial load forecasting is well explained in [40], [94]. A comparison and selection method of spatial forecasts is presented in [95]. [96] solved the problem of spatial load forecasting using local movement, and [97] combined trending and land-use based methods to develop a spatial load forecasting method. Although these sources explain the idea of spatial load forecasting, all of them only applied it to distribution systems. As the main interest of this research is to assist power planners, the focus is on developing spatial forecasting methods useful to generation and transmission planners. To get to that point the area under study (the SWIS) needs to be divided into different sub-regions proper for spatial studies.

Meshing is the act of dividing the area under study into small sections. Mesh generation is studied in detail in [98]. The main concern in [98]is to define mesh with optimal shapes to capture more information out of the system under study. For example, it provides ways to generate optimal meshes for finite element analysis purposes. Although the idea sounds very similar, the application of mesh generation is unique in this chapter and is different from the presentation in [98]. It has to be suitable for spatial studies of electricity load and power system planning. The author is not very concerned about the shape of the meshes being produced. The meshing method and resolution vary according to the utility type that is going to be planned. For instance, distribution utilities serve small loads distributed in small areas inside the service area. Regular meshing with high resolution can be very handy in this case. On the other hand, transmission and generation utilities cover a relatively large area where low resolution irregular meshing can be applicable.

Given that meshing is a vital prerequisite for spatial forecasting and planning steps, the area selection should be undertaken in a way that helps the future steps. Below is a list of important characteristics that help define a proper mesh in general power planning applications:

- Type of electrical load: Electricity load can be residential, commercial, industrial, or a combination of them. It will be a good idea to place the consumers of each type in one sub-section.
- Proximity: Consumers of close substations will be placed in one division.
- Population: The population concentration affects the load consumption in the area and it would be good to put the areas of similar population in one category.
- Average temperature and relative humidity: Because temperature and relative humidity are the most important weather inputs, which affect the electricity consumption, placing the areas of similar average temperature and relative humidity will help the spatial power system planning.
- Availability of renewable sources of energy: It is important to consider the availability of the renewable sources of energy in the meshing step. This decision will ease the studies around renewable energy integration into the grid.

Based on the requirements of the problem, meshing can be regular or irregular [94]. Regular meshing involves dividing the area under study into equal sub-sections, with each section presented by an equilateral convex polygon. Irregular meshing is used when the shape of the sub-sections is not the main requirement and the emphasis is more on putting the areas of the same characteristics in one group. See Figure 9 for more clarification.

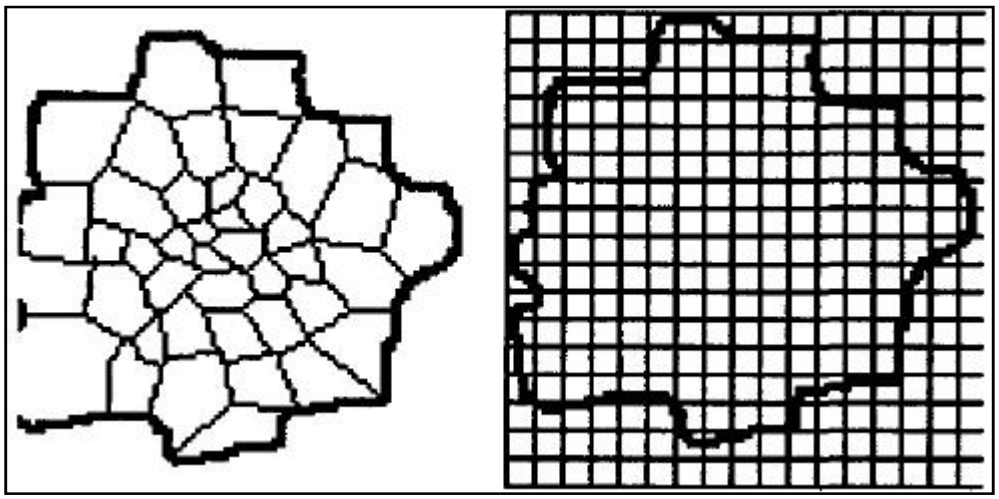


Figure 9: Irregular (left side) and regular (right side) meshing (extracted from [94]).

Appendix C presents some of Australia's distribution figures related to the above characteristics.

According to the mentioned characteristics, the SWIS can be divided into eight different sub-systems. The mentioned sub-systems are in accordance with the Western Power regions (The SWIS operator) as mentioned in [93]. Using the same regions as the ones used by the network operator makes the results of this work more understandable to people in the industry.

- CBD[20]: The CBD sub-system is the heart of the state's capital of Perth. With multi-storey office buildings, it is an ideal sub-system for commercial electricity load behaviours.
- Metro East: This sub-system is composed of the eastern part of the Perth metropolitan area and is mainly composed of residential buildings, making it an ideal location for residential load behaviours.
- Country Goldfields: Located in the eastern part of Western Australia and is mainly composed of mining companies, which makes it a perfect one for pure industrial studies.
- Metro North, Metro South, Country North, Country South, and Country East are other sub-systems of the SWIS. These regions are mixed regions with some dominated by industrial loads and others by residential loads.

To visualise these regions, the procedures shown in Figure 10 are followed and a computer program is coded to generate a mapping tool.

[20]Central Business District

Firstly, a map of the area is captured using Google map services and then, considering the above characteristics, the region is divided into eight different sub-systems.[21] Then the boundary coordinates are recorded and using a computer coded program the polygons are regenerated. The code can be found in section ix of Appendix D.

The generated meshes considering all the aforementioned characteristics are presented in Figure 11 and Figure 12. Figure 11 divides the SWIS area into five sub-sections, and Figure 12 magnifies the metro area and divides it into four smaller divisions. These maps will be used for presenting the calculated short-term and medium-term forecasts on a map. In Chapter 6, a potential application of this tool is shown.

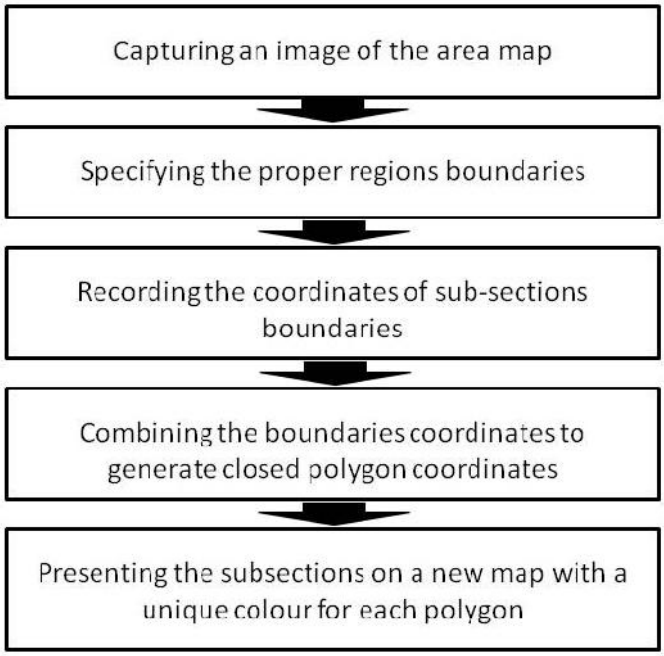


Figure 10: Meshing tool development steps.

3.5 CONCLUSION

This chapter described SWIS as the case study for this thesis. Different regions of SWIS are described and a meshing tool is developed to divide the area into sub-regions.

[21] The selected sub-systems are very similar to the ones defined by the local transmission system Operator Company, known as Western Power. The reason was that the output of this work can be readily applied to the existing system in the industry.

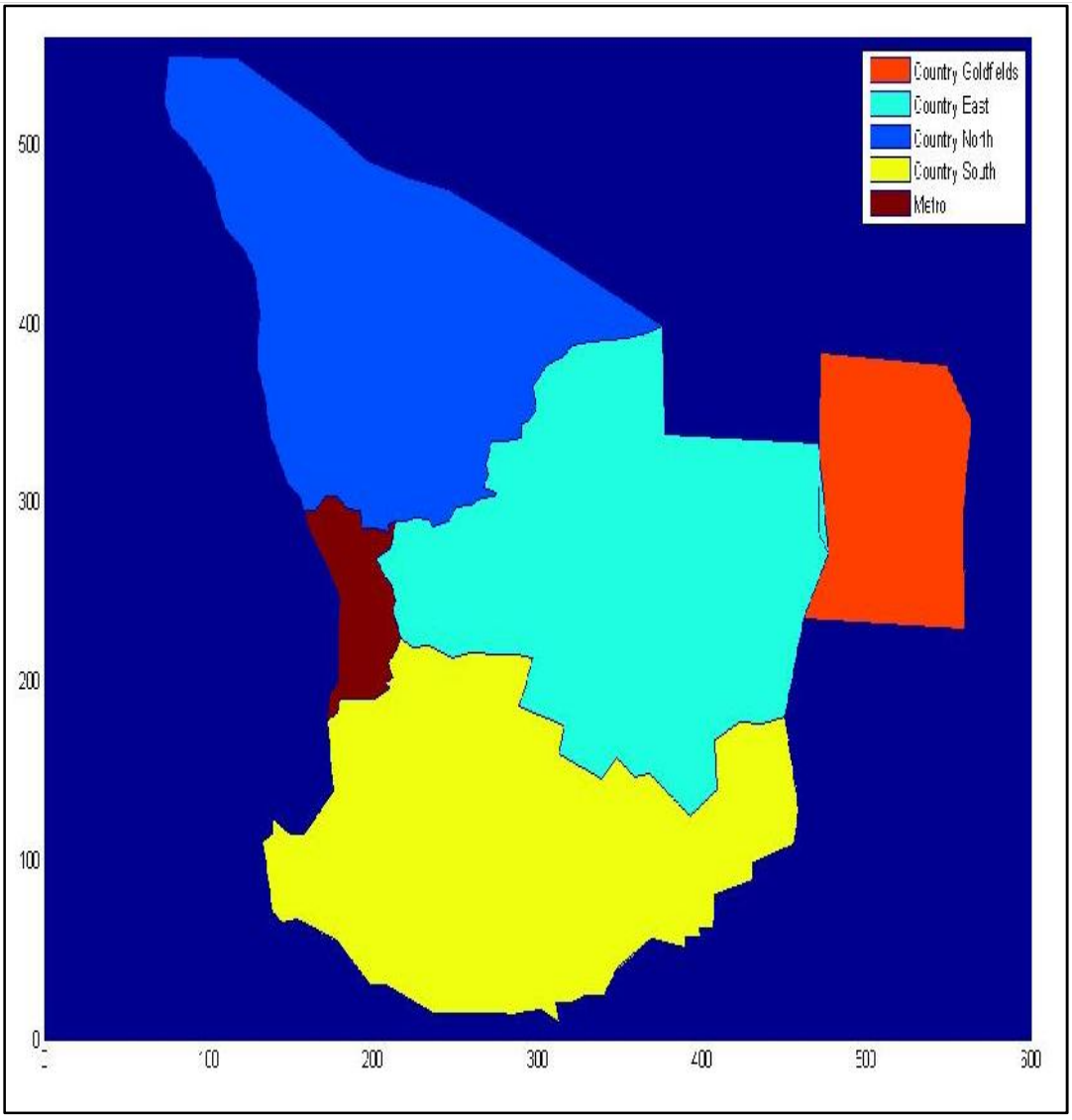


Figure 11: SWIS area meshing.

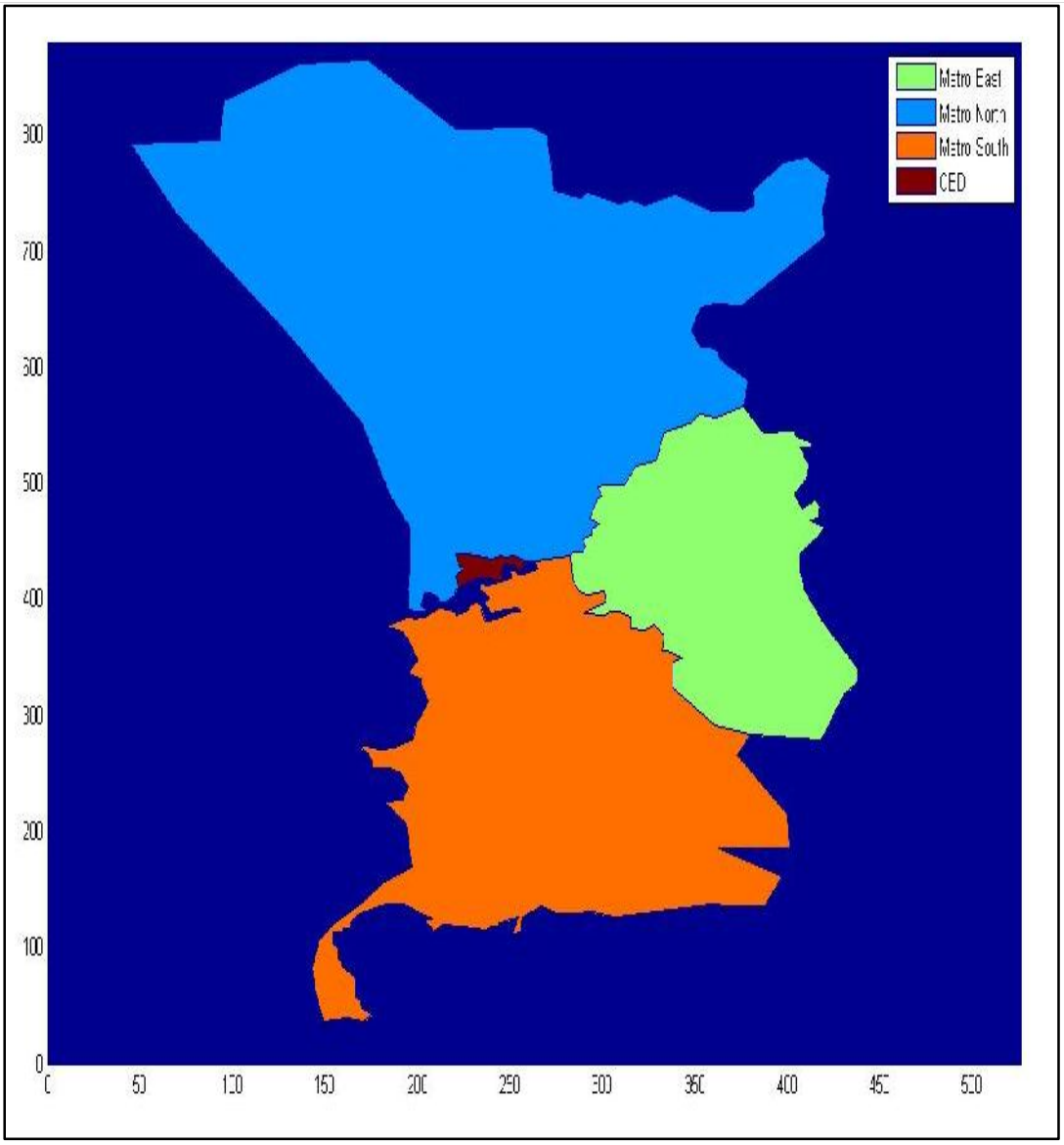


Figure 12: Metro area meshing.

Chapter

4 SHORT-TERM LOAD FORECASTER (CLASSIC APPROACH)

4.1 ABSTRACT

In this chapter, an innovative method for one day ahead and seven days ahead forecasts of electricity load is proposed. The proposed forecast method was tested on a residential load of SWIS. The case under study is tested under realistic conditions by considering minimum and maximum forecasts of temperature and relative humidity as available future inputs. Three different nonlinear approaches of neural networks, regression trees, and auto-regression with exogenous variables (ARX) were applied to fit a model with the mean absolute percentage error (MAPE). Artificial neural networks is chosen as representative for artificial intelligence-based methods, ARX is chosen as a representative of statistical-based methods. Regression trees as a regression method with if-then tree architecture represents a method that is both statistical-based and artificial intelligence-based. The results confirm improvement in all three training methods. As it was expected artificial intelligence-based model performs better than the statistical-based approach in short-term load forecasting applications. Regression trees with daily MAPE of 1.98% and weekly MAPE of 3.40% show the best performance. The obtained results confirm the functionality and high accuracy of this method. Limitations of this classic approach are also discussed, specifically, that the case study itself can dramatically affect the accuracy of the forecasting method.

4.2 INTRODUCTION

The complexities of contemporary electricity markets are enormous [99]. Electricity is traded based on bilateral contracts between energy providers and energy consumers. Electricity demand forecasting has a very significant role in present electricity markets as accurate load forecasting results in substantial financial savings and increased network reliability. Forecasts of a couple of hours to seven days ahead are known as short-term load forecasts (STLF). Applications of STLF include dispatching and commitment of generators, load shedding, and the determination of market prices.

Due to its essential role in the electricity market, many methods for generating STLF have been developed (for a review of previous methods, see [100]–[101] or literature review chapter). This chapter focuses on three methods: neural networks, regression trees, and ARX.

All these three methods are used to solve the load forecasting problem in the literature.[102] applied neural networks for hourly load forecasts. Applications of artificial neural networks in one day ahead load forecasts have been addressed in [103]–[104].An adaptive neural network approach was used for seven day ahead forecasts of electricity load in [105]. Regression trees were used to model a day ahead load in Spanish power systems [106] and have also been used in combination with fuzzy clustering [107]. An application of ARX in STLF is addressed in [108].

In this chapter, some modifications will be applied on these three methods to develop STLF of up to seven days ahead. Refer to 4.3for a full description of the steps taken. Considering the mentioned definitions of load forecasting horizons in Chapter 2, electricity load forecasts from a couple of hours ahead up to seven days ahead can be categorised as STLF. This chapter explores

the classic approach of load forecasting method for up to seven days ahead load forecasts with a very good level of accuracy and reliability. The reported results are slightly better than reported ones in the literature. This is because of additional data pre-processing works and model set up. The following section provides more details.

4.3 METHODOLOGY

The proposed methodology is a combination of data clustering, pattern recognition, and signal reconstruction techniques, and a training model to formulate the input-output relations. Three different nonlinear training methods of artificial neural networks (ANN), regression tree learning, and auto-regression with ARX have been used in this study. To construct more than one tree in regression tree method, techniques called ensemble methods can be used. The ensemble method can be bagging [109], random forest, boosted trees[110] and rotation forest. Bagging or bootstrap aggregating is a method used in machine learning algorithms to reduce variance and also help with over-fitting. Random forest is another method that can help with the problem of over-fitting. Due to its performance under classification noise, bagging has been selected for the construction of regression trees for ensemblesas reported in [111].

The main intent of this work is to perform short and medium-term load forecasts of up to one year ahead. GDP and population growth play critical load in long term electricity forecasts. However, for a time period of less than a year the change in parameters like population and GDP is minor and they hardly affect the electricity load. By using them, the complexity of model will increase and the accuracy will drop. As such, both GDP and population data are excluded from this work.

A summary of the proposed method is illustrated in Figure 13. This flowchart is gradually formed by the author during his move towards preparing high accuracy electricity forecasts. Here is how this chart is being concluded.

4.3.1 RESOLUTION ADJUSTMENT

As mentioned earlier, the resolution of electricity load data and weather information are not the same and the first step in data pre-processing will be data resolution adjustment. The resolution of provided temperature and relative humidity data were hourly. This is basically how the Australian Bureau of Meteorology records its data. However for short term load forecasts the accuracy will be higher if higher resolution is used. The resolution was changed to half hourly by replacing each half an hour data with the average of neighbouring points and using first order hold for each 30 minutes of data.[22]Refer to Appendix D for a computer code generated to adjust data resolution. Temperature and relative humidity are selected as the main two inputs to the training methods as they have shown much higher correlation values comparing to the other inputs studied.

[22]See Appendix D for the codes.

4.3.2 *DEALING WITH MISSING DATA AND OUTLIERS*

Scatter plots of the data sets show some visible outliers. Visible and non-visible outliers are removed in the next step. Investigating the data sets again show some missing data points. It could be because of sensor failure or data transmission failure. Those missing points are to be replaced by proper estimates.

As recommended by [112], one dimensional median filtering can be used to remove outliers. However, median filtering alone is not capable of removing all the outliers automatically. To capture normal outliers a short window should be applied. A long window should be applied after that to capture outliers in a row. Human supervision is also required to remove outliers. The human expert can change the window sizes and investigate the data graphs and Q-Q[23] plots to ensure that the outliers have been removed properly. He/she can also specify and remove the outliers by using Matlab brushing tool. The same tool will be used to remove load outliers. For a comprehensive study of Q-Q plots and outlier removal refer to Chapter 5. Missing data points are simply replaced by averaging neighbouring points. Refer to Appendix D for computer codes generated to estimate missing data points and outlier detection.

4.3.3 *CLUSTERING AND SIGNAL RECONSTRUCTION*

The forecasting resolutions of weather data are different. Although sometimes these forecasts are available for every three hour of the next week, to avoid the loss of generality only minimum and maximum values of temperature and humidity are considered to be available to this framework at the time of forecasting. While weather forecasters are able to provide seven days ahead forecasts for a range of variables in a limited resolution of time [113], forecasts beyond this horizon are not reliable because of the unpredictable nature of influential variables on weather systems. According to the Australian Bureau of Meteorology (BOM), these variables include: Surface pressure, Rainfall, Mean sea level pressure, Wind speed and direction, Geopotential height, Temperature, Relative humidity, Dew point, Combined sea and swell, Primary swell, Secondary Swell, Wind waves, and Wave period. This is why forecasted weather data cannot be directly used as an input for medium and long-term load forecasting applications. (For more information about this, refer to Chapter 6 where medium-term forecasting is being developed).

As stated by [29] the most important elements of weather for electricity demand forecasting are temperature and relative humidity. This can be also verified by a quick correlation test between load data and temperature and relative humidity. To be useful for practical applications, input data for load forecasting applications should be realistic and available at the time of running the framework. As discussed earlier, and not to lose the generality of this method we assume that only the maximum and minimum forecasts of temperature and relative humidity are available

[23] A Quantile vs. Quantile or Q-Q plot is a graph that shows the probability of two distributions at the same time. By using Q-Q plots similarities and differences of two different distributions can be investigated.

at the time of forecasts. Thus, only the minimum and maximum values of temperature and relative humidity for seven days are considered as the weather inputs of this framework.

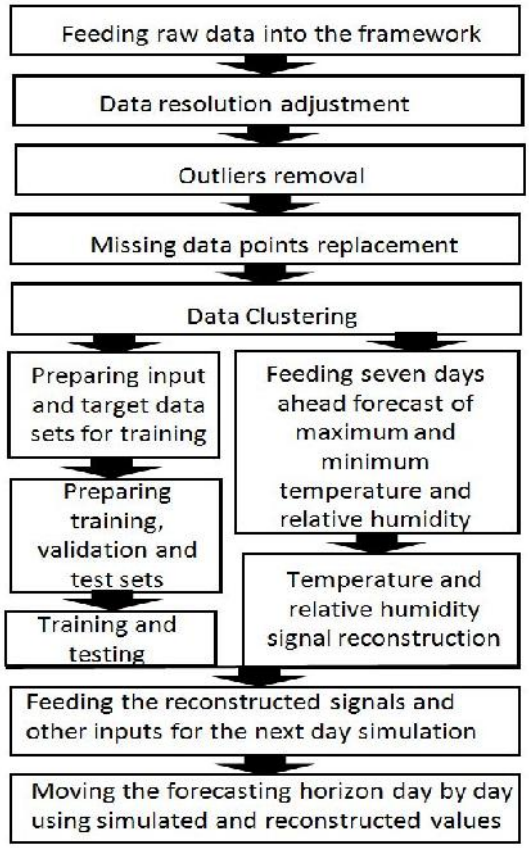


Figure 13: Proposed technique for seven-day ahead load forecaster.

To extract the weather distribution data out of available minimum and maximum forecasts, historical data of temperature and relative humidity are clustered[24]. Data sets of each cluster follow a similar pattern. By recognising these patterns and using the maximum and minimum forecasts of temperature and relative humidity, the cluster distribution signals are reconstructed. Half an hourly temperature and relative humidity information of the next seven days can be extracted from the reconstructed signals and be used for seven day ahead load forecasting purposes. See 4.4 for examples.

4.3.4 INPUT DATA

The raw input data is composed of seven days ahead minimum and maximum temperature, and relative humidity forecasts and historical load, temperature, and relative humidity data. After

[24]Clustering is used intensively in load forecasting applications to enable important information concerning seasonal weather variables and electricity consumption to be derived from the data set [135]–[136].

resolution adjustment, the removal of outliers and the replacement of missing data points with meaningful values, input matrixes were formed out of the available data. Preparing a set of input variables is a significant step in any training procedure as it can strongly impact on the accuracy of the method. Input data include either variables that correlate strongly with the output data, or variables that help to classify the other input variables. Introducing more input variables to the problem without high correlation with outputs is avoided to increase the accuracy of the model. Some of these inputs are indirectly used in the model. For example, by having an input as year number the model will be able to consider items like change of GDP and population indirectly.

After pre-processing and clustering the data, 13 sets of input variables were defined based on available temporal, load, and weather data as a feed for training models. The aforementioned inputs and a description for each are as follows:

1. Year: Load data of any example usually varies on an annual basis. Annual trends will vary dependent on the year as a result of population increase and industrial growth specific to the region within a particular time frame. By choosing a year as one separate input, these kinds of changes can be captured.
2. Month: Each month has its own specific patterns of temperature, relative humidity, wind and rain fall. These control variables introduce seasonality in the load data that needs to be captured.Considering months instead of seasons captures the seasonality of the data to a better degree.
3. Day of the week: Each day of the week has its own load pattern, as the needs of consumers differ on different days of the week. Although some weekdays have similar load patterns, each day of the week was individually treated in this method. The main reason is that similar days do not behave exactly the same. Therefore, placing them in less than seven groups would result in a loss of some vital information about the week days.
4. Hour of the day: Load pattern also varies from hour to hour during the day.
5. Temperature.
6. Relative humidity.
7. Previous day same hour load. This helps to capture the autocorrelations in load data.
8. Previous week same hour load. This helps to capture the autocorrelations in load data.
9. Holiday: Public holidays and weekends have some similarities in load patterns in the same way that weekdays have similar consumption patterns. Defining this variable and its combination with day of the week variable enables the consideration of different day types' load patterns. More information about the public holidays used in this research are detailed in Appendix B.
10. Average past 24-hour load: This variable highly affects each day's consumption. This is founded after running correlation tests between the average past 24-hour load and the next day load.
11. Average past seven days load: This variable shows a good correlation with load.
12. Summer temperature: This value helps the classification of temperature in hot days.
13. Winter temperature: This variable helps to distinguish the cold days' temperature.

4.3.5 *TRAINING AND SIMULATION*

Three different training methods of artificial neural networks, regression trees learning, and ARX were applied using selected sets of input and output data, as specified in the previous step. The optimal architecture for each method was achieved by testing different configurations and observing the output accuracy.

The architecture of the neural network used is feed forward back-propagation with 40 hidden layers and one output layer. This configuration had the best regression for both training sets and validation sets. Training performance is set on minimising the mean absolute error. Refer to Appendix D for more details on this configuration.

Decision trees architecture used considers 40 number of regression tree with 30 as the minimum leaf size.

In ARX model, the order of output polynomial is set as 18,and the order of thirteen output polynomials are set at 18,20,16,19,20,16,20,20,20,19,20,19,20.

After testing the trained model, the next step was simulating the results. Using selected sets of inputs, single day ahead forecasts for any single day can be achieved. For seven days ahead forecasts because the input data for each day is not available some of the inputs need to be forecasted as the new input for a new day up to seven days.

Refer to Sample Codes for setting up neural networks, regression trees and calculating the error.

4.4 CASE STUDY

A pure residential area, East Perth metropolitan area, was selected as the case study to test this method. A pure residential load was chosen to capture all the behaviours of households in that area without the additional complications of industrial load regularities[25]. Additionally, a pure residential load model enables testing under the worst case. That's because residential consumers' behaviour is very nonlinear. They are not following a daily schedule like commercial or industrial consumers. Therefore, when the same method is applied on a combination of industrial load and residential load[26], the output accuracy will be increased. The region consists of one 6.6 kV and six 22 kV distribution substations and the total of 17 transformers.

As mentioned in section 4.3, historical data of load and weather of the region under study are required. The load information of each individual transformer was extracted from the database of Western Power, the company responsible for building, maintenance, and operation of Western Australia's South West electricity grid, known as the SWIS. The author approached Western Power (the SWIS network operator) in 2010. After several communications and phone discussions they mentioned that the load information is not readily available and data

[25]Industrial loads have a completely different behaviourcompared to residential data. They are mainly influenced by the situation of economy and factory work schedule. That's why it is better to study them separately. For more information on load system behaviours refer to Chapter 5.
[26]Commercial load behaviours are somewhere between residential and industrial loads. For more information on load system behaviours refer to Chapter 5 system-based load forecasting approach.

extraction is required. The author then moved to Western Power office and spent three months to extract all the demand load data of the SWIS. The data was extracted from two data bases called PI and Saffaire and included fifteen years of load data of all the substations of the region. The data resolution was half an hourly and the size of the files was extremely big. Weather data of the same period of time and locations was provided by Australian Bureau of Meteorology. The resolution of weather data was hourly.

The load data were combined to determine the electricity load of the East Perth region. The specifications of input data to the short-term forecasting framework are presented in Table 2.

Table 2: Raw data specifications.

Data	Unit	Resolution	Start date	End date
Load	MW	Half hourly	01-Jan-1995	01-Jan-2011
Temperature	°C	Hourly	01-Jan-1995	01-Jan-2011
Relative Humidity	%	Hourly	01-Jan-1995	01-Jan-2011

Figure 14 shows the scatter plot of electricity consumption versus temperature over 15 years of observation. It can be used to observe the correlation between temperature and load changes. In air-conditioning science around $20 - 23^{o}C$ is known as the comfort region. Most of humans do not use means of heating or air-conditioning in these temperatures where the temperature-load correlation is close to zero. At higher temperatures there is a positive correlation with load and for lower temperatures the correlation is negative. This can be visually confirmed just by having a close look on the scatter plot. When the temperature increases above 23°C the load does also increase. In other words a positive correlation exists between load data and temperature in hot seasons. Alternatively when the temperature decreases below 20°C the load data again increases. In mathematical-terms this is a negative correlation which happens in cold seasons. This confirms the fact that people in Perth mainly use electricity for cooling and heating purposes. Unlike the region under study, for regions where gas heaters are widely used, the correlation of load with low temperatures is close to zero. An example of such region is presented in[99].

After clustering the data, more information can be captured from temperature and load relations of each cluster. Figure 15 and Figure 16 show daily temperature and load distribution of a sample cluster. In our study each cluster is a day of a certain month. For example Wednesdays of April make one cluster. Using half hourly resolution each cluster will consist of 24 (number of hours in a day) x 2(number of half hours in an hour) x 4.5 (number of weeks in a month) x 15 (number of years) = 3240 data points. A regular pattern can be easily seen in both of them. With the help of clustering, weather signals can be accurately reconstructed and also proper inputs for training models can be generated.

By recognising these patterns and using the maximum and minimum forecasts of temperature and relative humidity, the cluster distribution signals are reconstructed. Half an hourly

temperature and relative humidity information of the next seven days can be extracted from the reconstructed signals and be used for seven day ahead load forecasting purposes. For each day of the 7 day week there are 68 similar days to search into. The search is being done by looking for closest maximum and minimum temperature to the forecasted one.

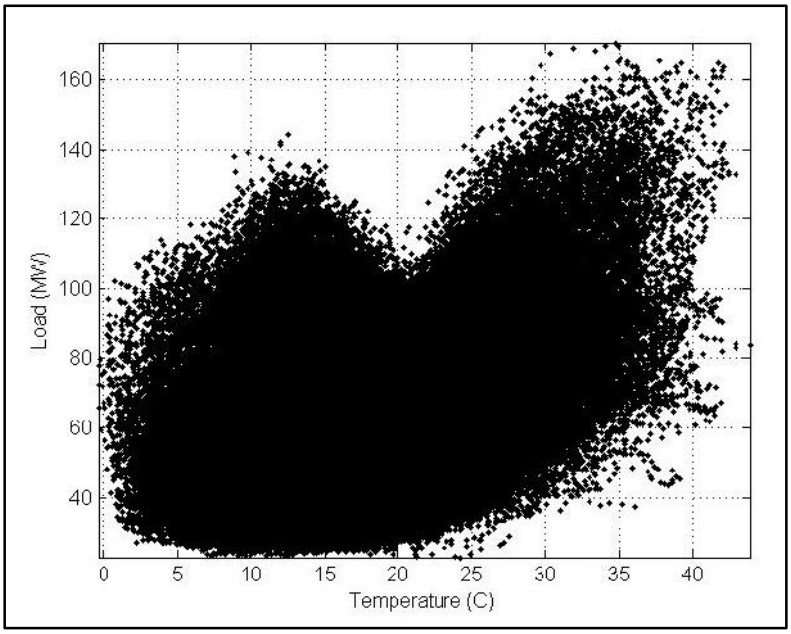


Figure 14: Half hourly electrical load consumption (MW) versus temperature data (degrees Celsius) from 15 years of observations for East Perth metropolitan area.

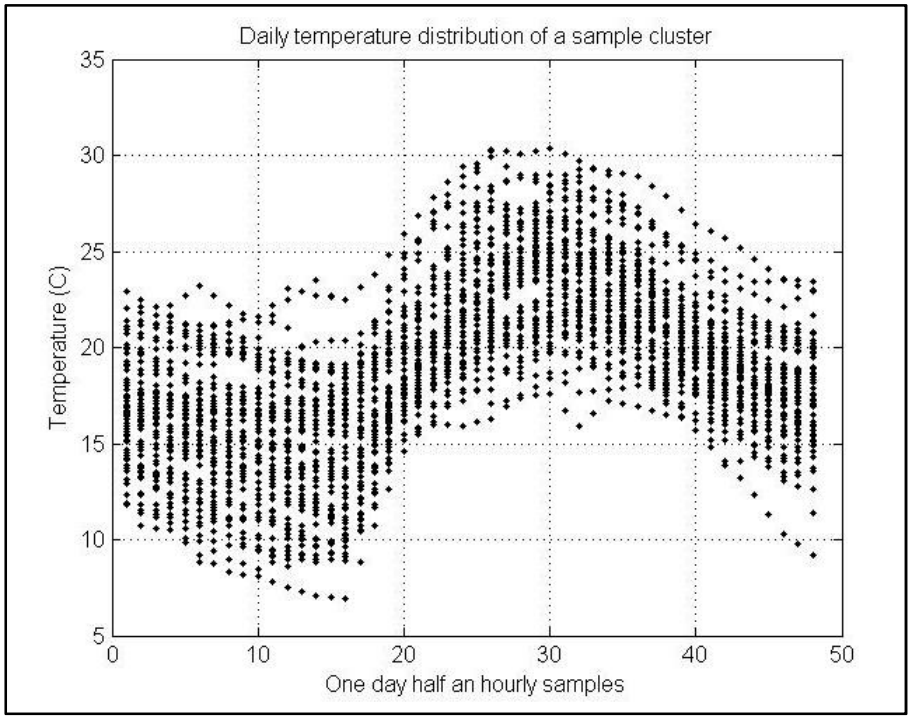


Figure 15: Daily temperature distribution of a sample cluster.

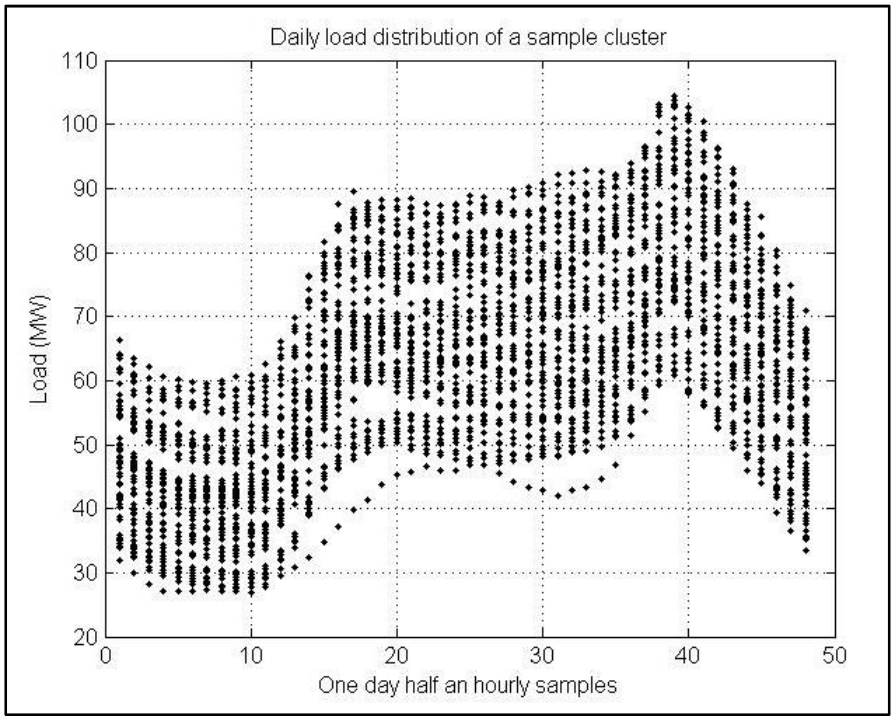


Figure 16: Daily load distribution of a sample cluster.

4.5 FINAL RESULTS AND DISCUSSIONS

The models have been trained with 14 years of data from January 1995 to January 2010. The testing period is 2010. The one day ahead and seven days ahead forecasts of a sample week in April 2010, are respectively presented in Figure 17 and Figure 19. Their residuals are shown in Figure 18 and Figure 20.

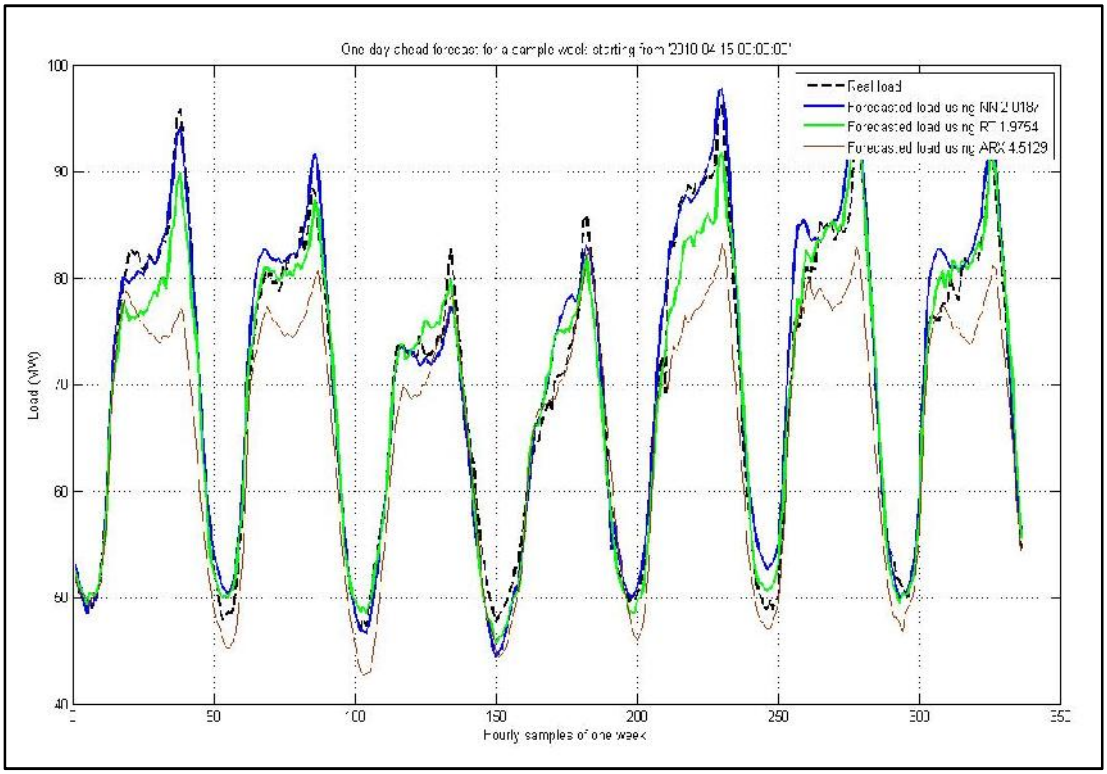


Figure 17: One day ahead forecasts of a sample week using three methods.

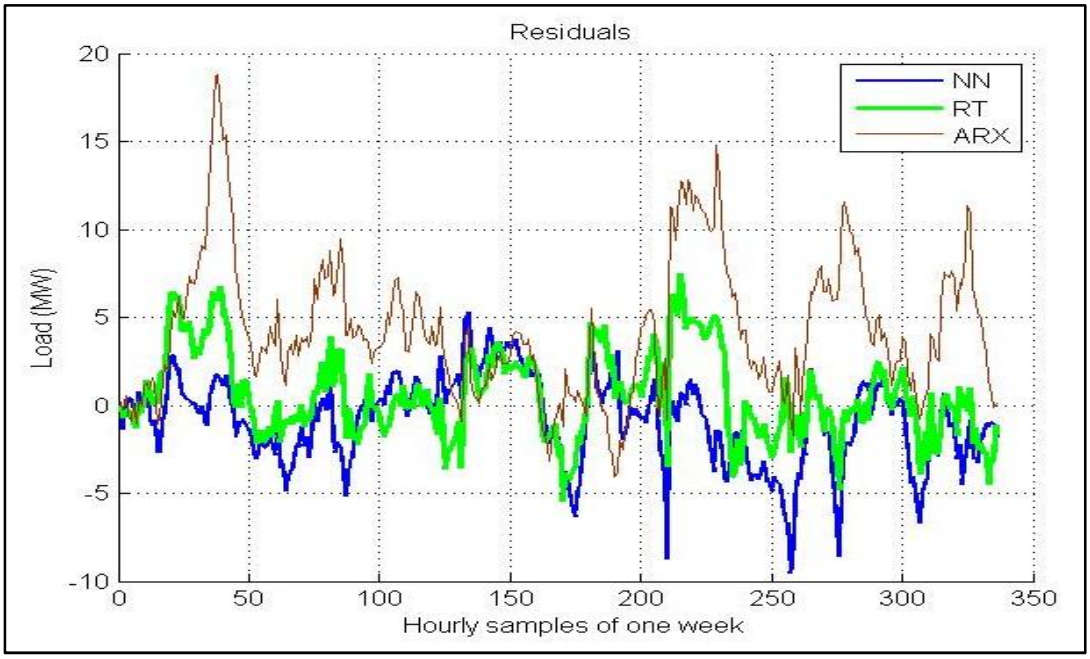


Figure 18: Residuals of one day ahead forecasts for a sample week.

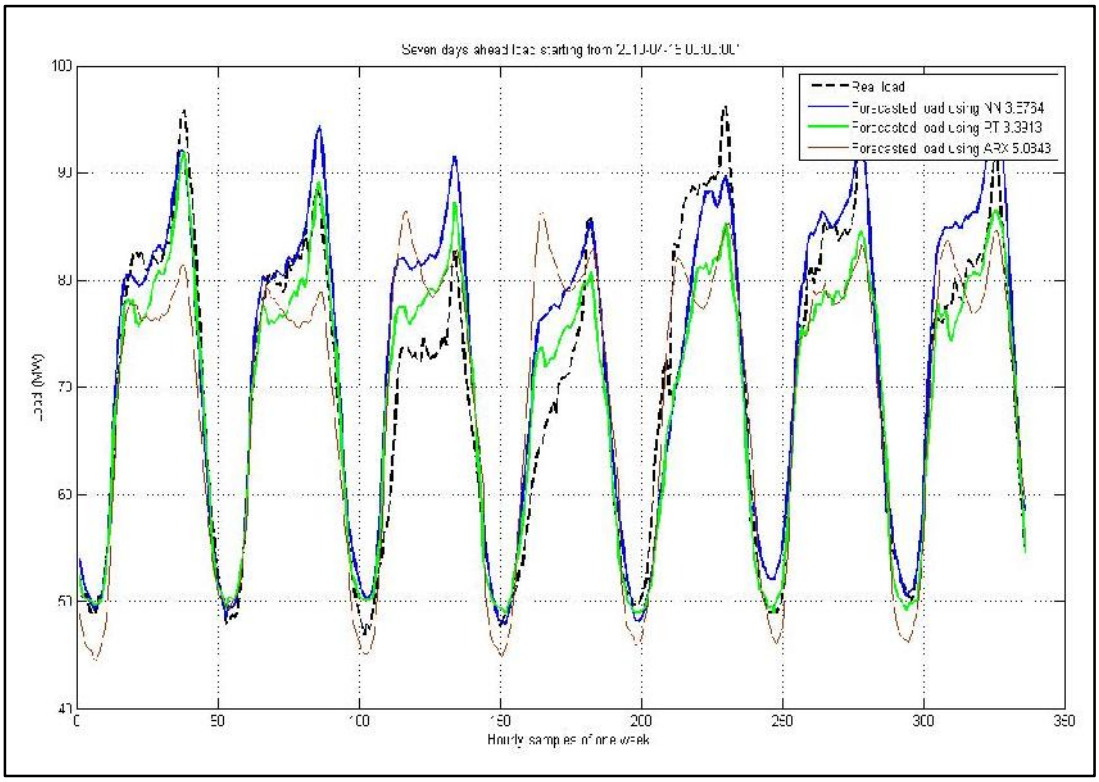


Figure 19: Seven days ahead forecasts of a sample week using three methods.

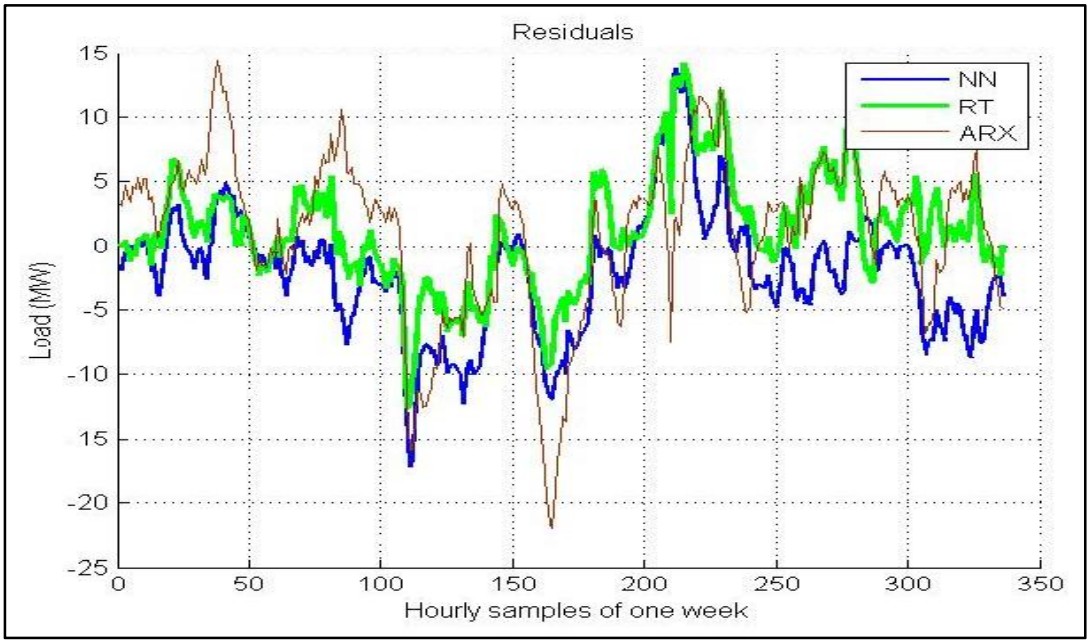


Figure 20: Residuals of seven days ahead forecasts for a sample week.

The mean absolute percentage error (MAPE) was calculated as the comparing criterion. MAPE can assign a number to the error of each method. MAPE has become somewhat of a standard in load forecasting applications. For few other works using MAPE to report the model accuracy the reader can refer to[114], [115] and [116]. The following is the formula that was used:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{l_{Ai} - l_{Fi}}{l_{Ai}}\right| \qquad (4\text{-}1)$$

Where:

l_{Ai}	Actual electrical load in MW
l_{Fi}	Forecasted electrical load in MW
n	Number of forecasted data points

Table 3 shows the comparison between the three different methods that were used. Daily MAPE is the mean absolute percentage of each method for one day ahead forecasts of a week (see Figure 17) and weekly MAPE presents the accuracy of each method for seven days ahead forecasts (see Figure 19).

Table 3: Comparing the error and speed of applied training algorithms.

Training algorithm	Training speed	Daily MAPE	Weekly MAPE
Neural networks	Slow	2.02 %	3.88 %
Bagging regression trees	Fast	1.98 %	3.39 %
ARX	Fast	4.51 %	5.08 %

[117] states that it is relatively an easy task to get to the MAPE of 10% in load forecasting application. But each percent lower MAPE in the forecast of electricity can potentially save up to £10 million in operating costs per year as investigated by [118].

Comparing the achieved results from neural networks, bagging regression trees, and ARX respectively with conventional methods of neural networks[114], regression trees, and time series [119] show improvement in accuracy for all of the tree training methods that have been used. The improvement could be because of the more detailed data pre-processing methods introduced in this chapter. Among all of the applied methods bagging regression trees shows the best results using the presented methodology.

The improvement in the forecasting accuracy comes from, pre-processing of raw input data as stated in 4.3.1 and 4.3.2, size of the data set extracted, resolution of the data set extracted, clustering and signal reconstruction techniques as described in 4.3.3 and the 13 sets of input variables defined in 4.3.4.

4.6 Conclusion

In this chapter, the problem of seven days ahead load forecasting has been solved. The proposed framework generates the seven days ahead forecasts based on a stable next day forecast and the seven day forecasts of maximum and minimum temperature and relative humidity. A pure residential area has been selected as the case study to test the model under the worst conditions. For the training step of the proposed solution three different algorithms have been applied to investigate their ability to generate the best forecast. Regression trees show the best performance among the three methods. The accuracy of regression trees in this application is slightly better than neural networks and the training speed is noticeably higher. Although the achieved accuracy is good for seven days ahead forecasts, better accuracy is expected for case studies with some level of industrial load involved. The reason is less nonlinearity and randomness in behaviour.

4.7 Challenging the classic approach

Although the results are sound, this method is not necessarily the optimal solution to be used in any load system. The problem with a classic approach is the limited generalisability of the single

case study to other load profiles, and in particular the variability resulting from the unique system behaviour of the selected case study. The following chapter presents different case studies, and based on the load determination criterion, a new rule will be presented to help the load analysts select the best forecasting method based on electricity load distributions. As MAPE may not be the best method of comparing different forecasting methods, a modified version of MAPE called MMAPE will be presented in the next chapter.

Chapter

5 SHORT-TERM LOAD FORECASTER (SYSTEM BASED APPROACH)

5.1 ABSTRACT

An innovative method of forecasting electricity load was introduced in the previous chapter. The forecast accuracy of the classic approach is satisfactory compared to the conventional methods of short-term load forecasting (STLF). However, in a classic approach, there are some unseen items that can dramatically affect the forecast accuracy and the reliable application of the method on other load systems. The major missing part in the classic approach is not considering the behaviours and characteristics of the system under study. In other words, in classic approach, a method of forecasting gets developed and applied to all load forecasting problems. In this work the reader will see that a method that is efficient on a case study may not be as efficient on another one. Another important missing part is the detailed study of input variables, such as, temperature and relative humidity, and looking at their effects on system behaviours. This chapter will examine those unseen parts and will introduce a system based approach that can be adapted to any existing load system to produce more accurate and reliable STLF.

5.2 INTRODUCTION

To better understand the deficiencies of the classic approach, this chapter will examine the behaviours of eight regions in SWIS with different characteristics. The regions are: Metro North, Metro East, Metro South, CBD, Country North, Country South, Country East, and Country Goldfields. The load pattern of each region is different. Some are purely residential, some are dominantly industrial, CBD is a mixture of commercial and residential and some are mixture of all three. As the data extraction and pre-processing steps are similar to the previous chapter for all the regions, only a summary of the procedure is presented below. The focus of a system based approach is different from a classic approach and gives the forecaster a tool to select the best forecasting method by analysing the system load data and temperature sensitivity.

5.3 WEATHER DATA

The weather data comprises the temperature and relative humidity of seven locations in the South West of Western Australia. The chosen weather stations data can be used for eight regions selected earlier. For some cases the weather data of two or three stations are averaged to achieve the weather parameters of the region. The weather data of these locationswas provided by the Australian Bureau of Meteorology (BOM)[113] upon the author'srequest. The data resolution is hourly and is from January 1995 to January 2011. However, according to the BOM reports, very few locations have a complete record of weather data. There are various reasons for missing data. The site may have been closed, reopened, and upgraded to a full weather site or downgraded to a rainfall only station during its existence, leading to some missing data. There might also be missing data due to a damaged instrument or the absence or

illness of an observer. Some outliers are also visible because of machine faults or small noises.Table 4 contains the raw weather data provided by the BOM.

Table 4: SWIS raw weather data.

Location	Variables	Start-date	Finish-date	Total number of observations
Perth	T and RH	01-01-1995	01-01-2011	280,514
Mandurah	T and RH	01-01-1995	01-01-2011	280,514
Bunbury	T and RH	22-04-1999	01-01-2011	206,508
Kalgoorile	T and RH	01-01-1995	01-01-2011	280,514
Geraldton	T and RH	01-01-1995	01-01-2011	280,514
Albany	T and RH	01-01-1995	01-01-2011	280,514

A number of data analysis and adjustment steps needed to be performed on the weather data to make them ready for the spatial STLF application. Here is how the weather data is mapped to each region.

- CBD: Perth weather data is used for this region.
- Metro North: Perth weather data is used for this region.
- Metro South: Average of Perth and Mandurah weather data is used for this region.
- Metro East: Perth weather data is used for this region.
- Country North: Geraldton weather data is used for Country North.
- Country South: Average of Albany and Bunbury weather data is used.
- Country East: Average of Kalgoorlie and Perth weather data is used for Country East.
- Country Goldfields: Kalgoorlie weather data is used for this region.

5.3.1 RESOLUTION ADJUSTMENT

As mentioned earlier, the resolution of electricity load data and weather information are not the same and the first step in data pre-processing will be data resolution adjustment. The resolution of provided temperature and relative humidity data were hourly. This is basically how the Australian Bureau of Meteorology records its data. The resolution was changed to half hourly by replacing each half an hour data with the average of neighbouring points and using first order hold for each 30 minutes of data.[27]

5.3.2 DEALING WITH MISSING DATA AND OUTLIERS

As recommended by [112], one dimensional median filtering can be used to remove outliers. The median filter considers each data point in a dataset and looks at its nearby points to decide whether or not it is representative of its surroundings. If a large change is detected, the data point will be replaced by the median of neighbouring values. However, median filtering alone is not capable of removing all the outliers automatically. To capture normal outliers, a short

[27]See Appendix D for the codes.

window should be applied. A long window should be applied after that to capture outliers in a row. Human supervision is also required to remove outliers. The human expert can change the window sizes and investigate the data graphs and Q-Q (Quantile vs. Quanltile) plots to ensure that the outliers have been removed properly.

For single outliers that are detectable by human eye and are clearly standing out of the data set, the human supervisor can specify and remove them by using tools like Matlab brushing tool. The removed single points will be treated as missing data points later on. As suggested by [112], a straightforward technique to estimate and replace the missing data points is with the average of neighbouring points.

The Q-Q plot is a graphical technique for determining if two data sets come from populations with a common distribution. By using Q-Q plots similarities and differences of two different distributions can be investigated. If the data sets are following the same distribution, the Q-Q plot will follow a 45 degree straight line. For more detailed information on Q-Q plots and their uses in probability theory [120] can be read.

Several items can be investigated by looking at Q-Q plot of two data sets. For the purpose of this work we are only interested in finding answers for below questions:

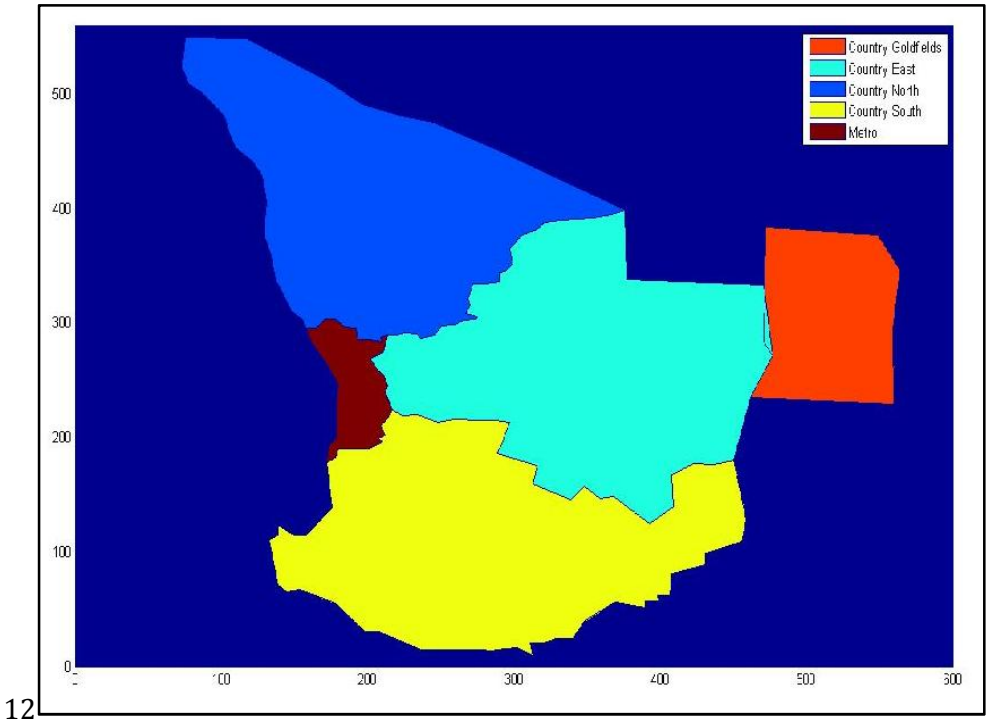
- Are there any outliers present? (Has the removal of outliers been efficient and successful?)
- Do two data sets come from similar distribution?

To examine the presence of outliers in a data set the Q-Q of the data set against normal (Gaussian) distribution can be drawn. Normal distribution is used because as stated in [121], physical quantities that are the result of many independent processes (such as measurement error) often have distributions very similar to normal. Single points away from the 45 degree line show the presence of outliers in a data set. Refer to 5.4.1 to see how normal Q-Q plots are used in this work to test the data sets for presence of outliers.

Later on in the thesis under section 5.5.2, Q-Q plots of other distributions are tested to find invaluable information about the behaviour of load data. To date, such behavioural analysis of load data has not been addressed by any other sources available in the literature. And the author can be considered as the first people who applied this.

5.4 LOAD DATA

Electricity load data of 440 transformers was extracted from the database of Western Power. The data resolution is half hourly and is from January 1995 to January 2011. By finding the locations of transformers on the SWIS map and aggregating the load data, different regions load could be achieved. The regions are the same as the eight regions described in 3.4. Country North, Country South, Country East and Country Goldfields are presented in Figure 21.Figure 22 breaks down the Metro Region into four smaller regions called CBD, Metro North, Metro East



12

Figure 21: SWIS area regions on map (country side regions).

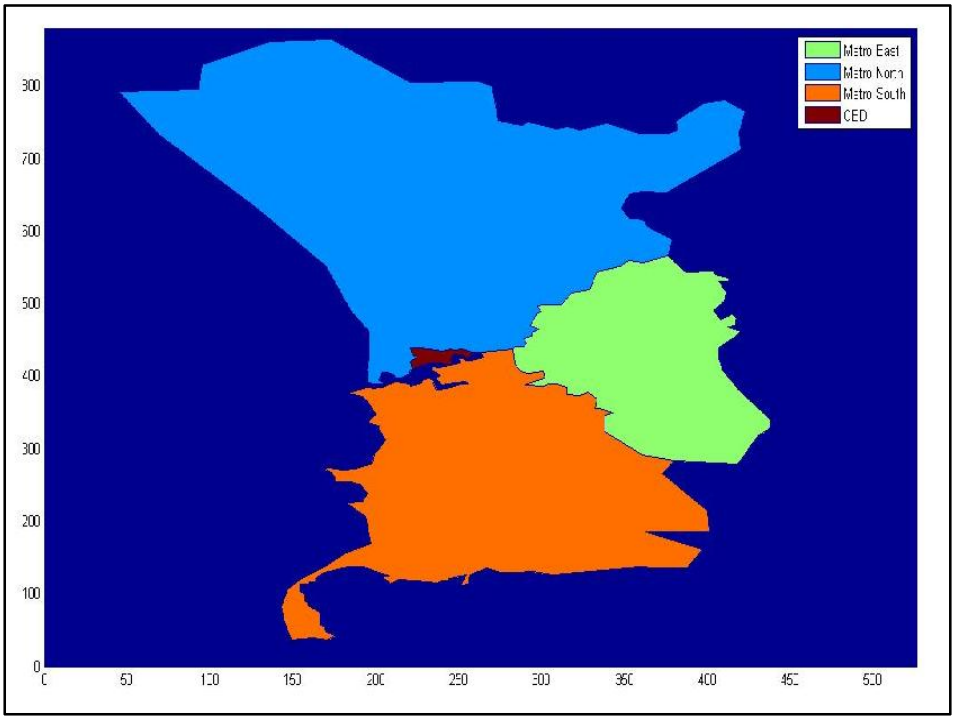


Figure 22: Metro area regions on map.

and Metro South. The recorded load data is full of outliers and missing values. To provide a better view of the raw data, scatter plots of half hourly load samples for each region are presented below (Figure 23to Figure 30). To plot the following figures, data cells with error messages and no values should be replaced by NAN values.

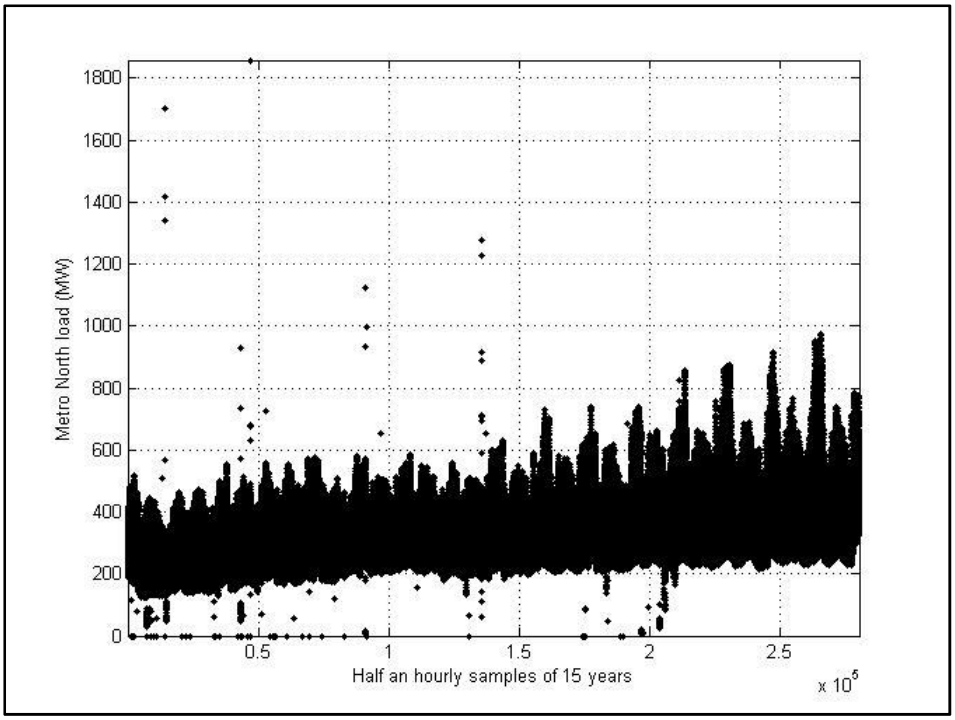


Figure 23: Metro North electricity load data from 15 years of observation.

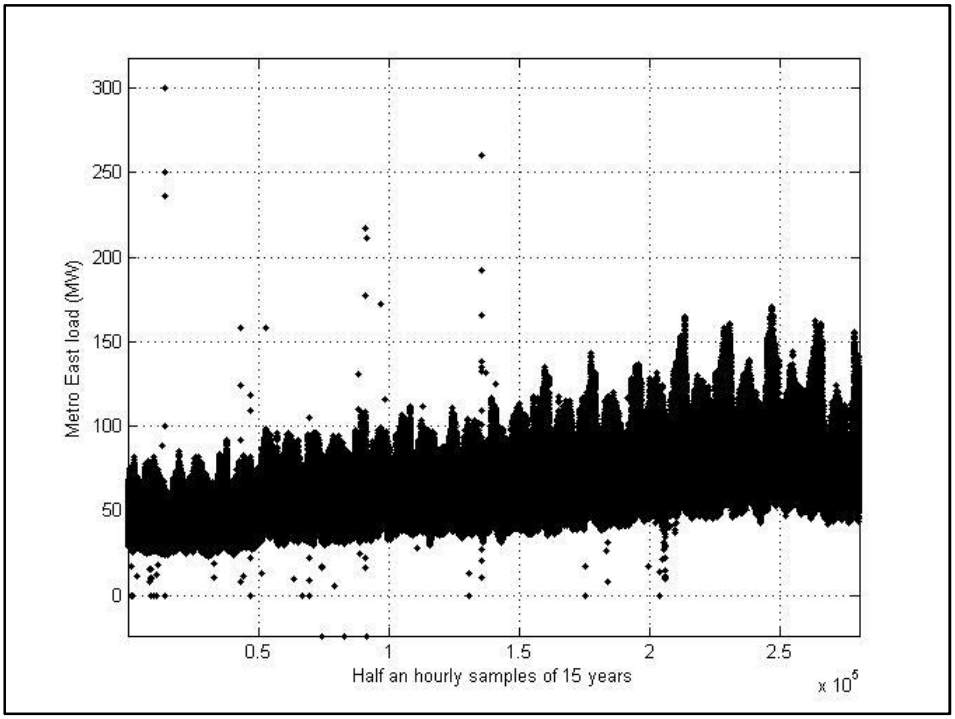


Figure 24: Metro East electricity load data from 15 years of observation.

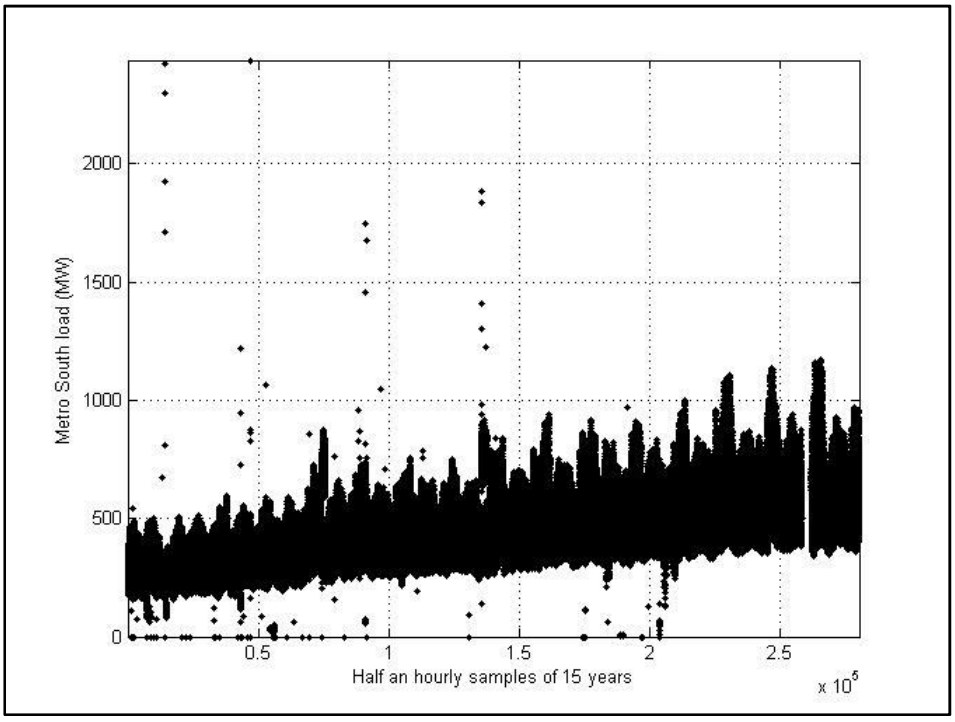


Figure 25: Metro South electricity load data from 15 years of observation.

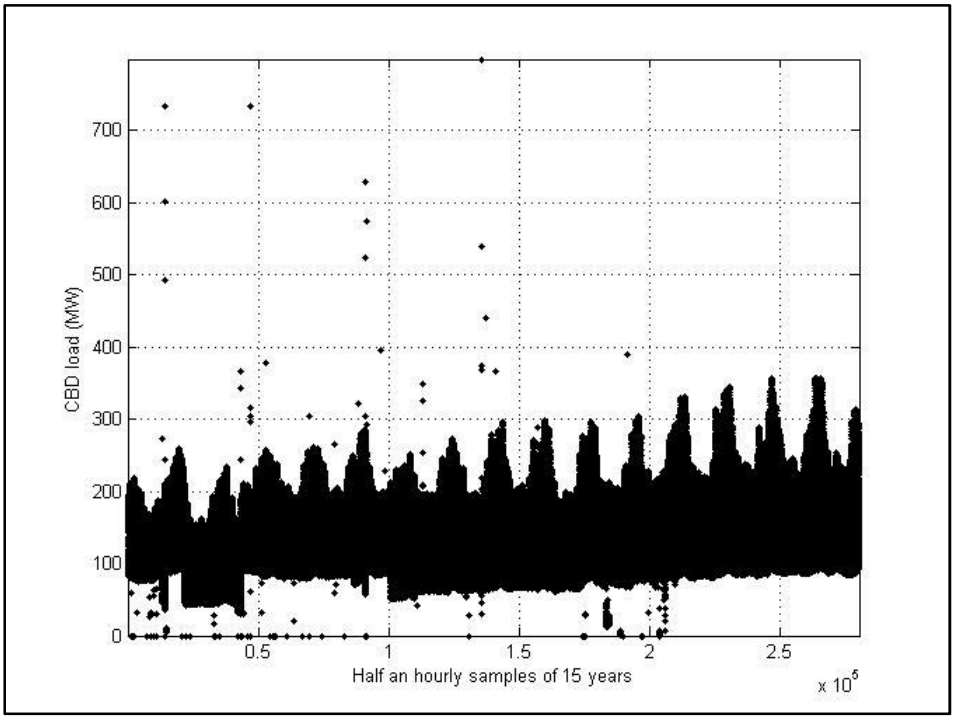


Figure 26: CBD electricity load data from 15 years of observation.

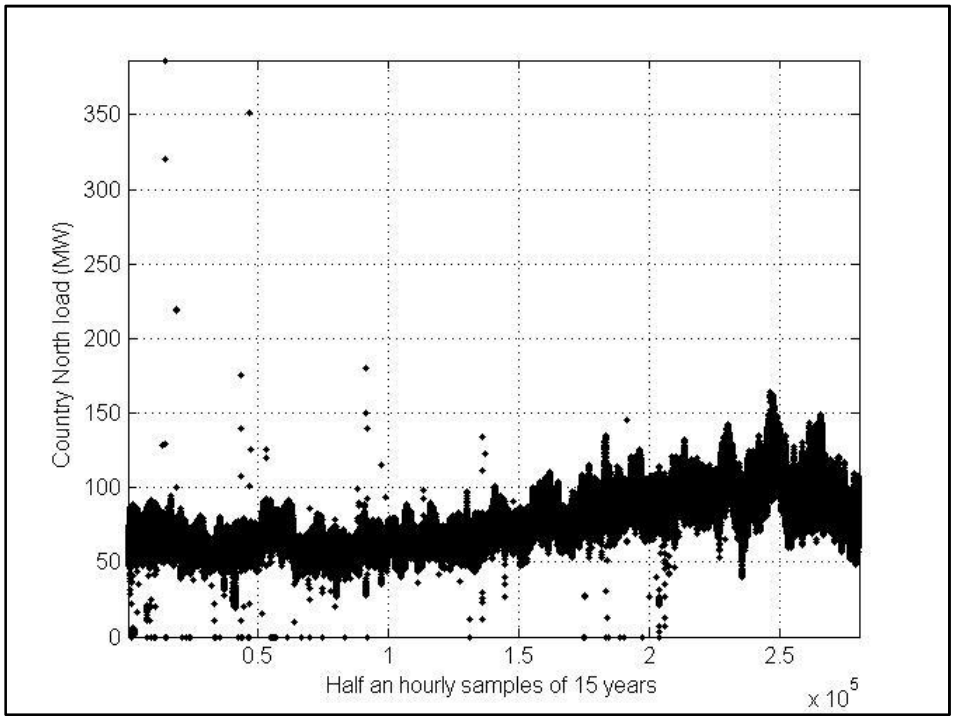


Figure 27: Country North electricity load data from 15 years of observation.

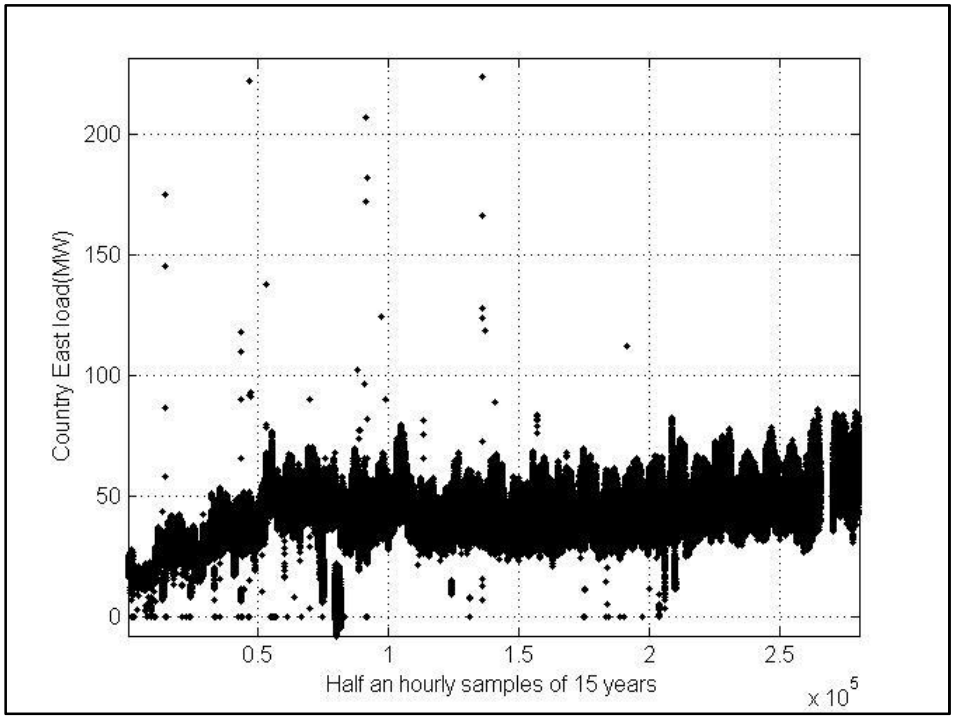


Figure 28: Country East electricity load data from 15 years of observation.

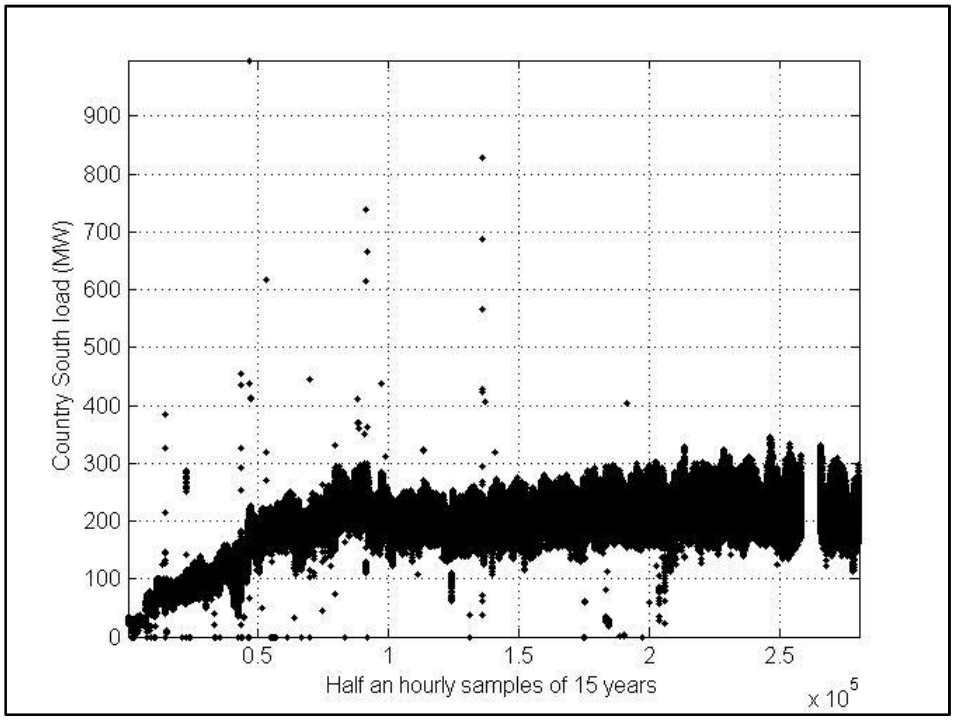


Figure 29: Country South electricity load data from 15 years of observation.

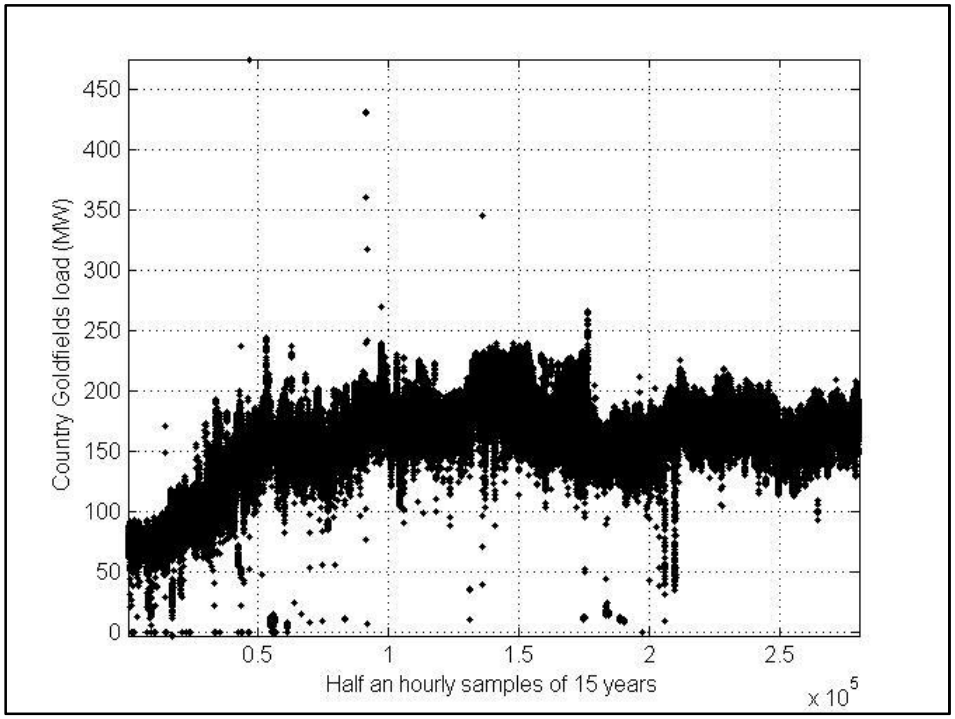


Figure 30: Country Goldfields electricity load data from 15 years of observation.

5.4.1 DEALING WITH MISSING DATA POINTS AND OUTLIERS

For single outliers that are detectable by human eye and are clearly standing out of the data set, the human supervisor can specify and remove them by using tools like Matlab brushing tool. The removed single points will be treated as missing data points later on. As suggested by [112], a straightforward technique to estimate and replace the missing data points is with the average of neighbouring points. Based on the idea that the load of each hour of a day is similar to the load of a similar hour of a similar day of the neighbouring weeks, the missing values have been interpolated using the four neighbouring weeks.[28] The author has come into the same conclusion as discussed under data clustering section 4.3.3.

After applying this method on the available data sets the Figure 31to Figure 38were obtained. By comparing these figures with the previous eight figures (Figure 23 to Figure 30), the efficiency of this method to remove outliers can be confirmed. Although the removals of some outliers are pretty obvious just by visually comparing the graphs, Normal Q-Q plots of the new sets and raw sets will be presented in the next section to see the efficiency of this outlier removal tennique.

[28]See Appendix D for the codes.

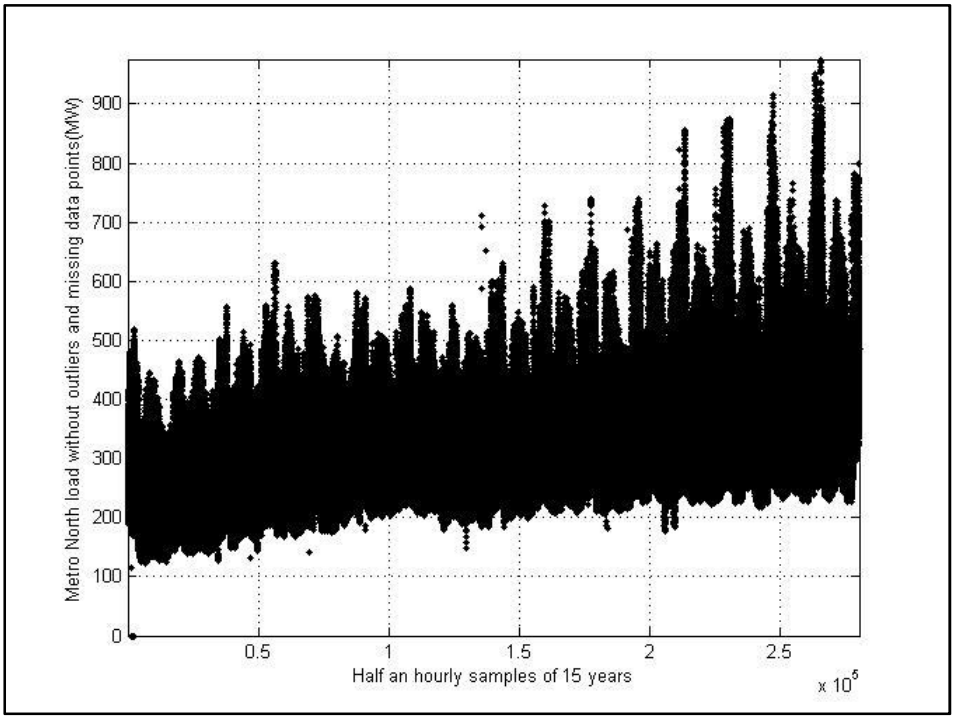


Figure 31: Metro North electricity load data with removed outliers and estimated missing data points.

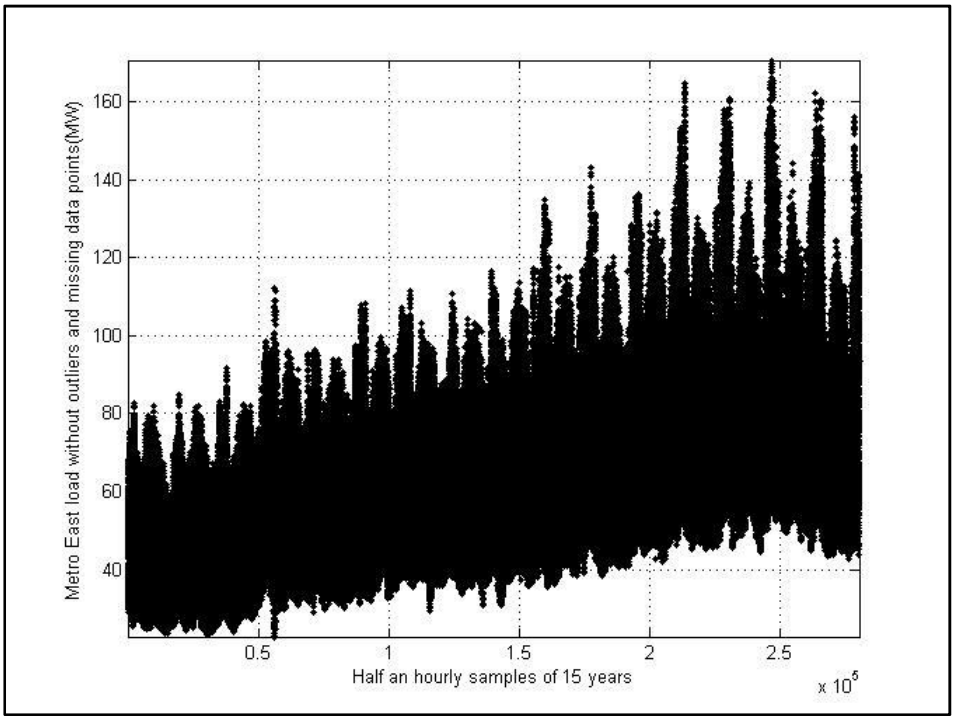


Figure 32: Metro East electricity load data with removed outliers and estimated missing data points.

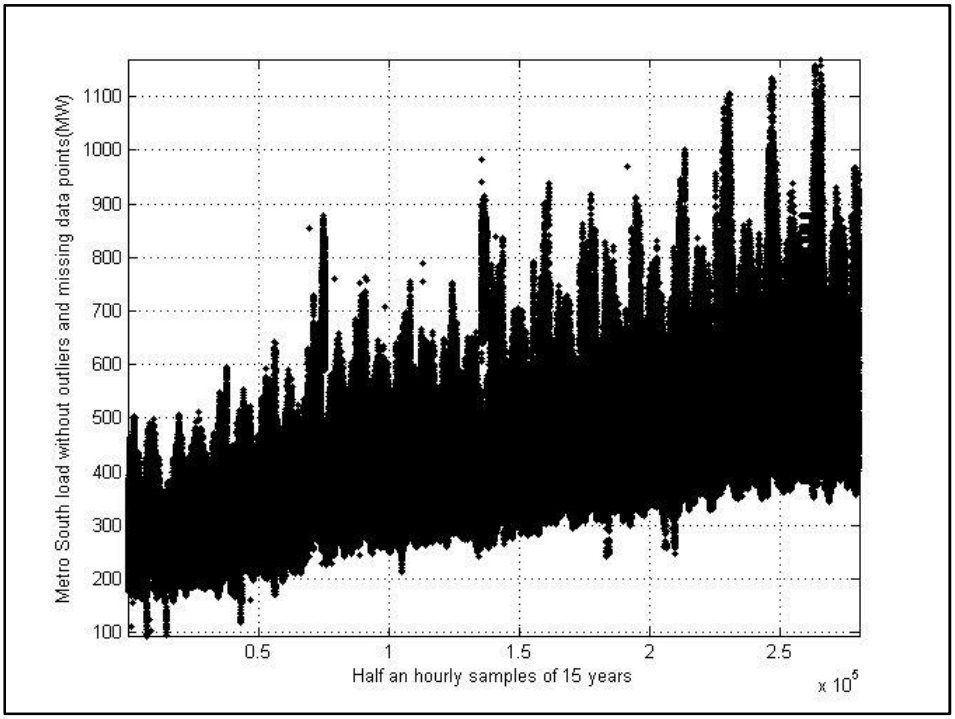


Figure 33: Metro South electricity load data with removed outliers and estimated missing data points.

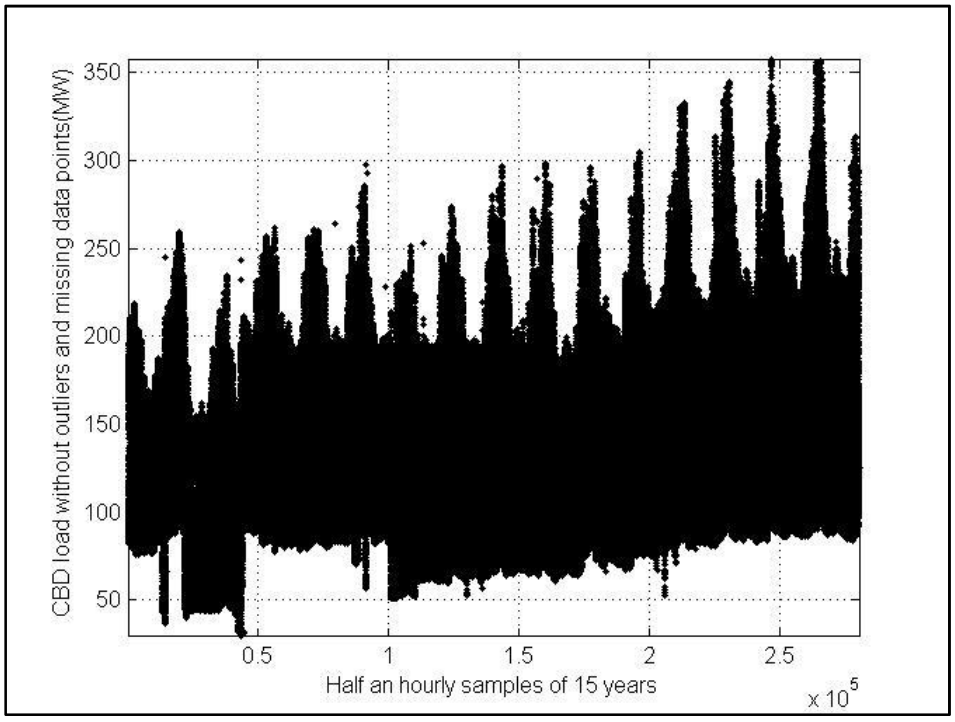


Figure 34: CBD electricity load data with removed outliers and estimated missing data points.

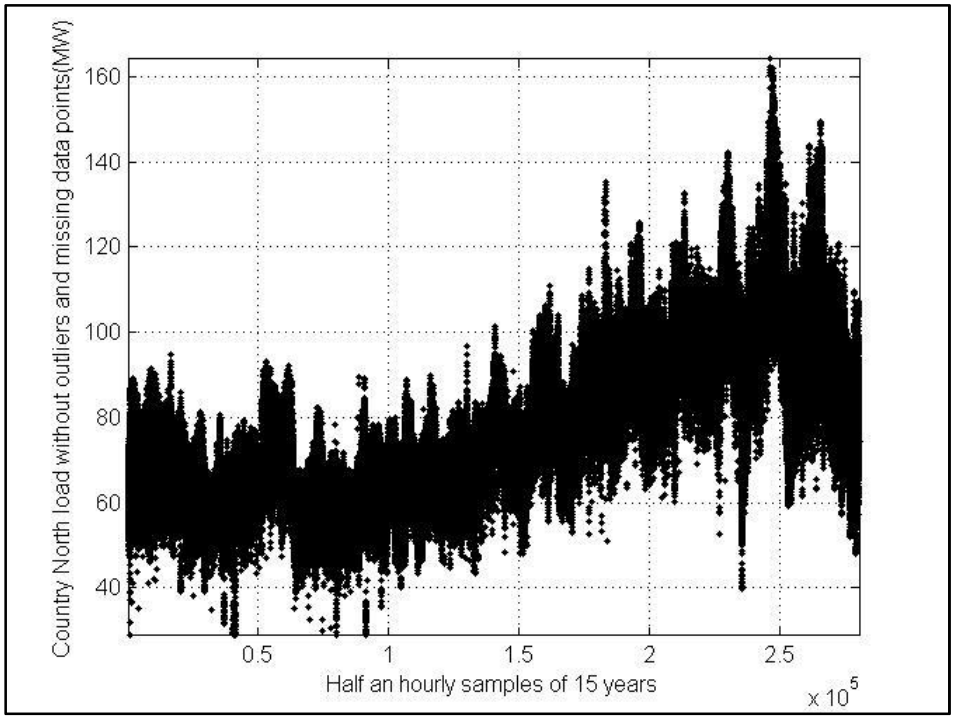


Figure 35: Country North electricity load data with removed outliers and estimated missing data points.

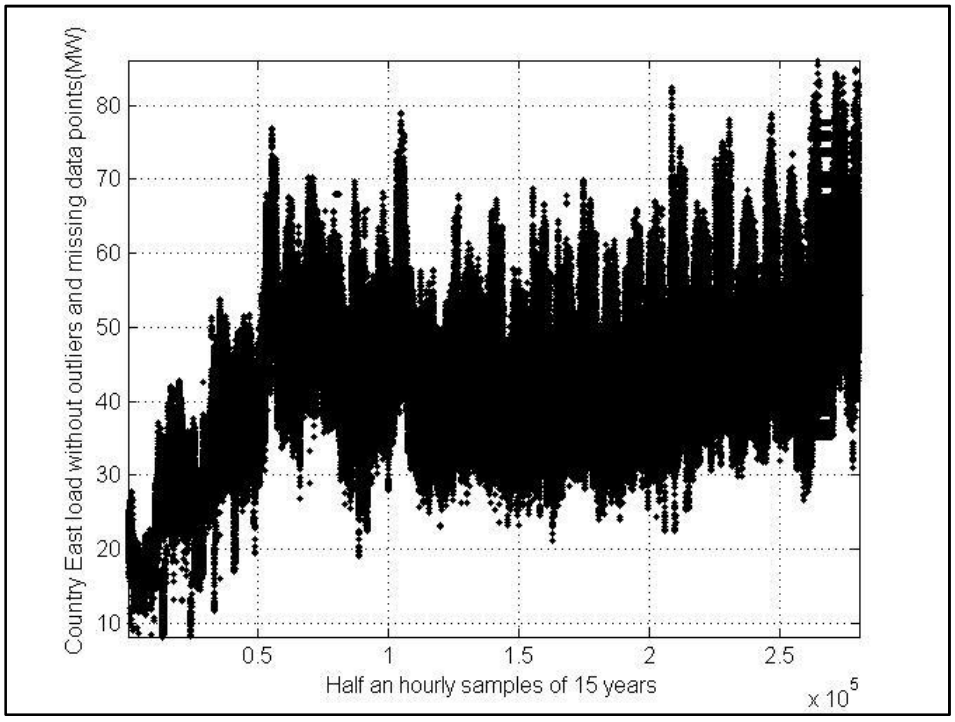


Figure 36: Country East electricity load data with removed outliers and estimated missing data points.

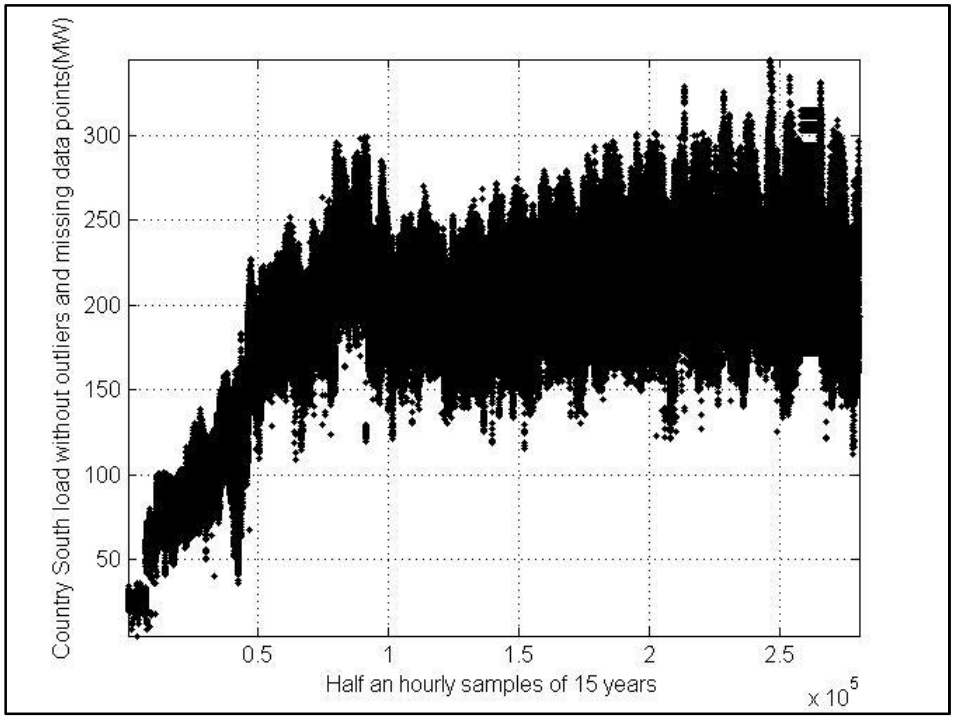


Figure 37: Country South electricity load data with removed outliers and estimated missing data points.

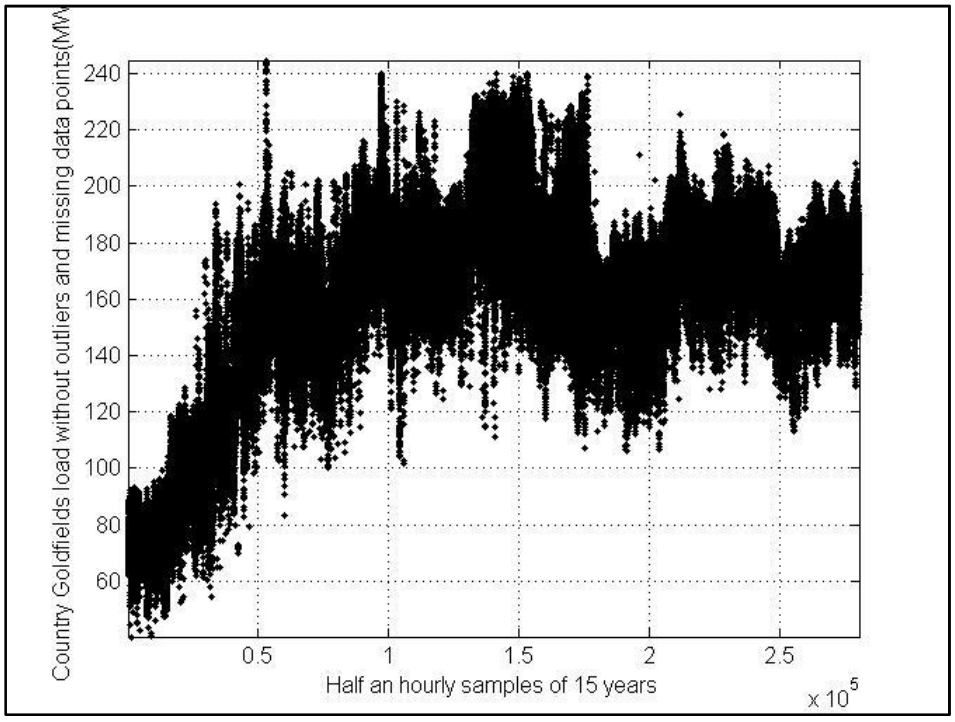


Figure 38: Country Goldfields electricity load data with removed outliers and estimated missing data points.

To better evaluate the capability of this method to remove outliers, standard normal Q-Q plots for all eight regions are presented below. Normal distribution is used because as stated in [121], physical quantities that are the result of many independent processes (such as measurement error) often have distributions very similar to normal. Single points away from the 45 degree line show the presence of outliers in a data set. In all cases, the left panel shows the quantiles of raw data versus standard normal, and the right panel represents the same Q-Q plot after the outliers have been removed. The outliers can be easily identified by comparing the two panels.

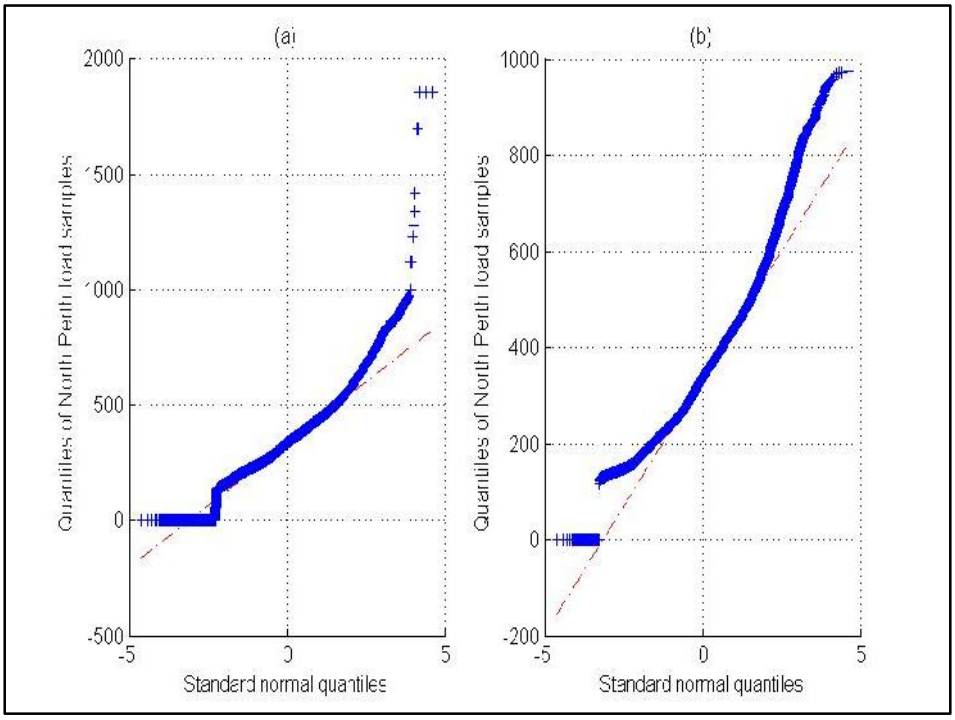


Figure 39: (a) Quantiles of Metro North raw load versus the quantiles of standard normal; (b) Quantiles of Metro North load with outliers and missing data points removed versus the quantiles of standard normal.

Figure 39, Figure 40, Figure 41, Figure 42, Figure 43, Figure 45 and Figure 46: By looking at both normal Q-Q plots it can be concluded that the data distribution does not exactly follow the distribution of standard normal. If that was the case the majority of the blue dots should be located on the 45 degree line showed as a dashed line. However, for this specific test we are only after investigating the efficiency of the outlier removal methods used.
Figure 44: By looking at both normal Q-Q plots it can be concluded that the data distribution is very similar to the distribution of standard normal. If that was the case the majority of the blue dots should be located on the 45 degree line showed as a dashed line.

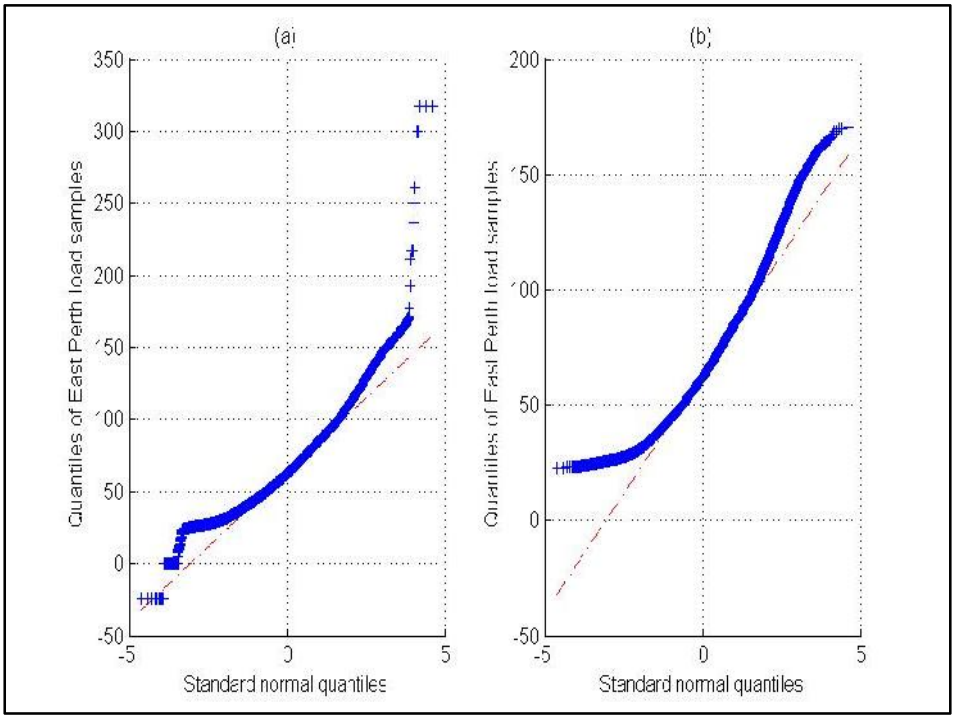


Figure 40: (a) Quantiles of Metro East raw load versus the quantiles of standard normal; (b) Quantiles of Metro East load with outliers and missing data points removed versus the quantiles of standard normal.

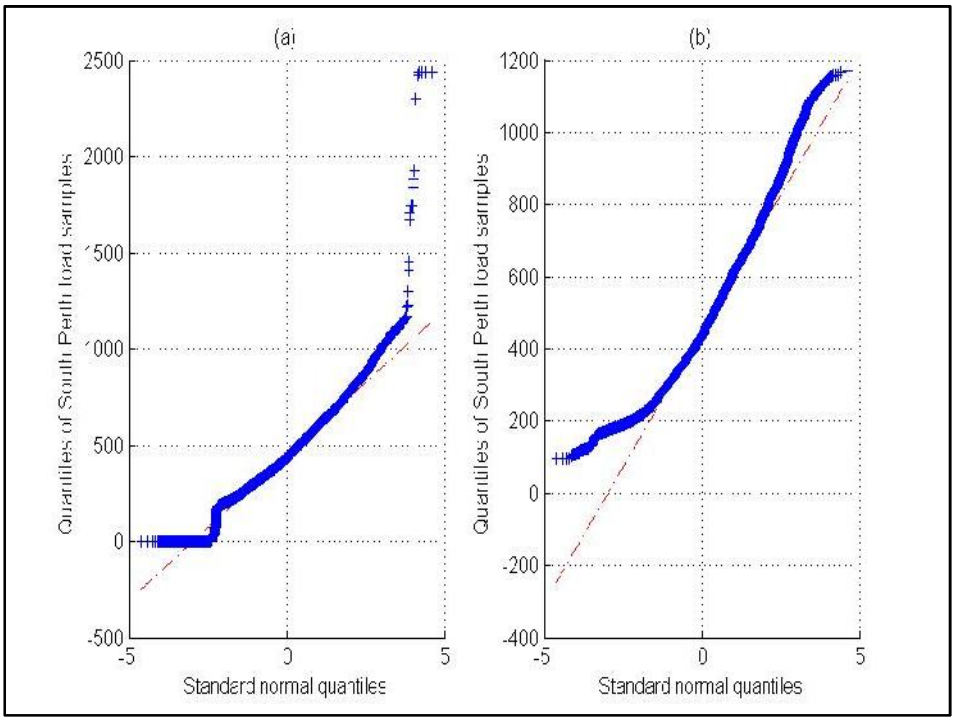


Figure 41: (a) Quantiles of Metro South raw load versus the quantiles of standard normal; (b) Quantiles of Metro South load with outliers and missing data points removed versus the quantiles of standard normal.

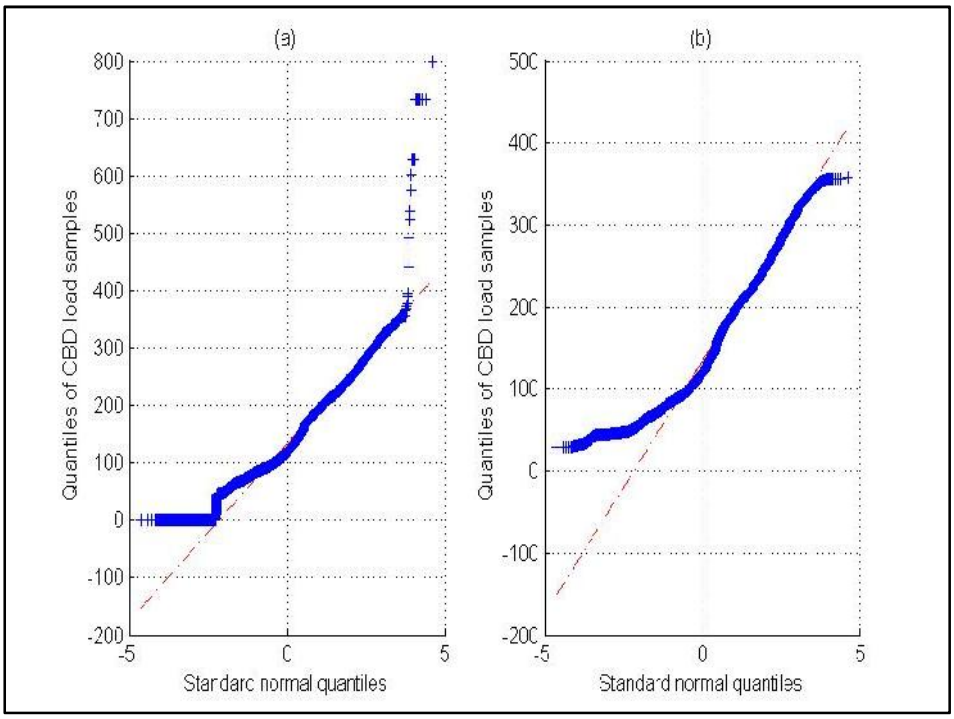


Figure 42: (a) Quantiles of CBD raw load versus the quantiles of standard normal; (b) Quantiles of CBD load with outliers and missing data points removed versus the quantiles of standard normal.

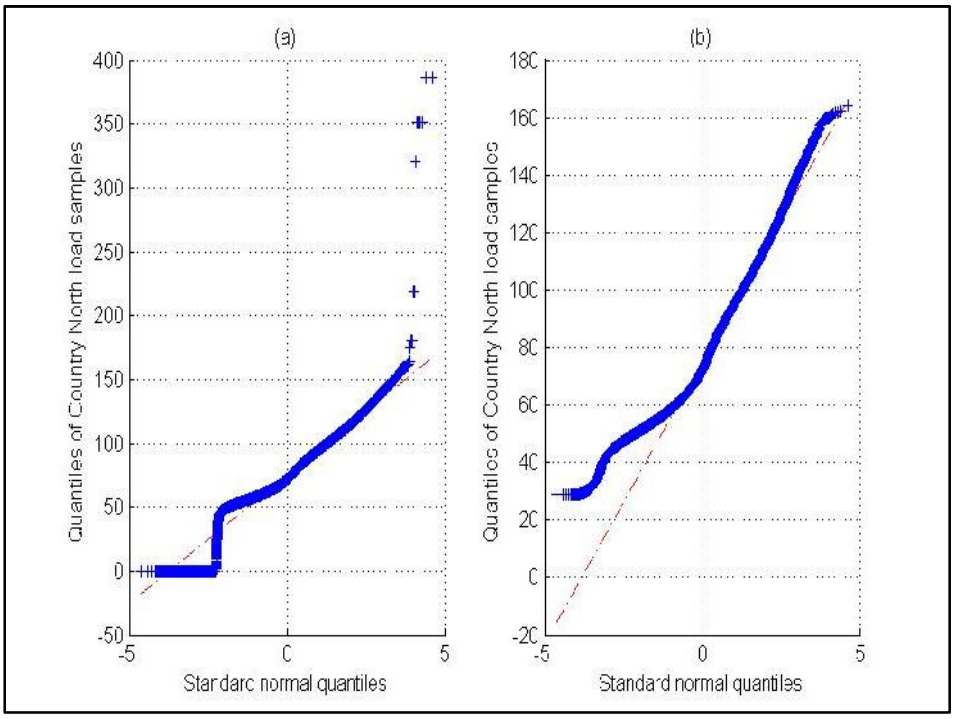


Figure 43: (a) Quantiles of Country North raw load versus the quantiles of standard normal; (b) Quantiles of Country North load with outliers and missing data points removed versus the quantiles of standard normal.

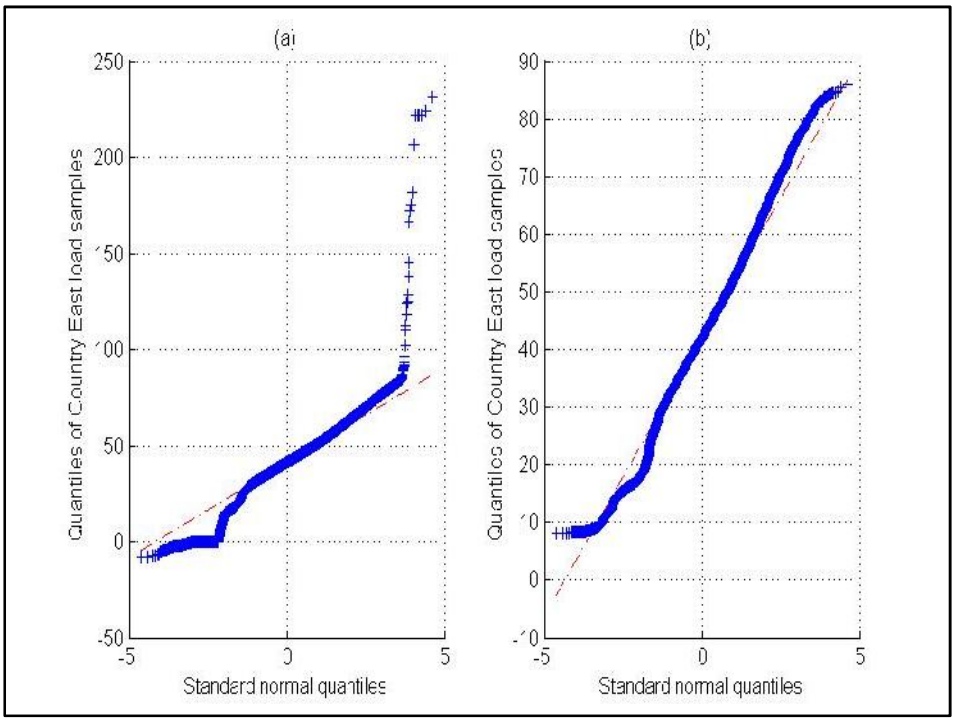


Figure 44: (a) Quantiles of Country East raw load versus the quantiles of standard normal; (b) Quantiles of Country East load with outliers and missing data points removed versus the quantiles of standard normal.

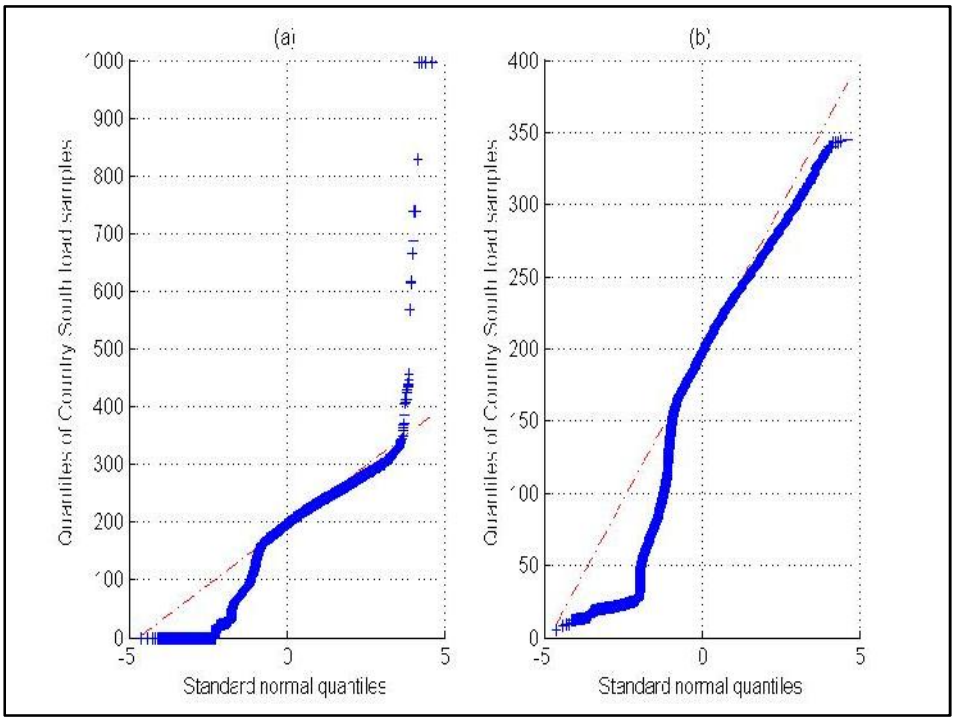


Figure 45: (a) Quantiles of Country South raw load versus the quantiles of standard normal; (b) Quantiles of Country South load with outliers and missing data points removed versus the quantiles of standard normal.

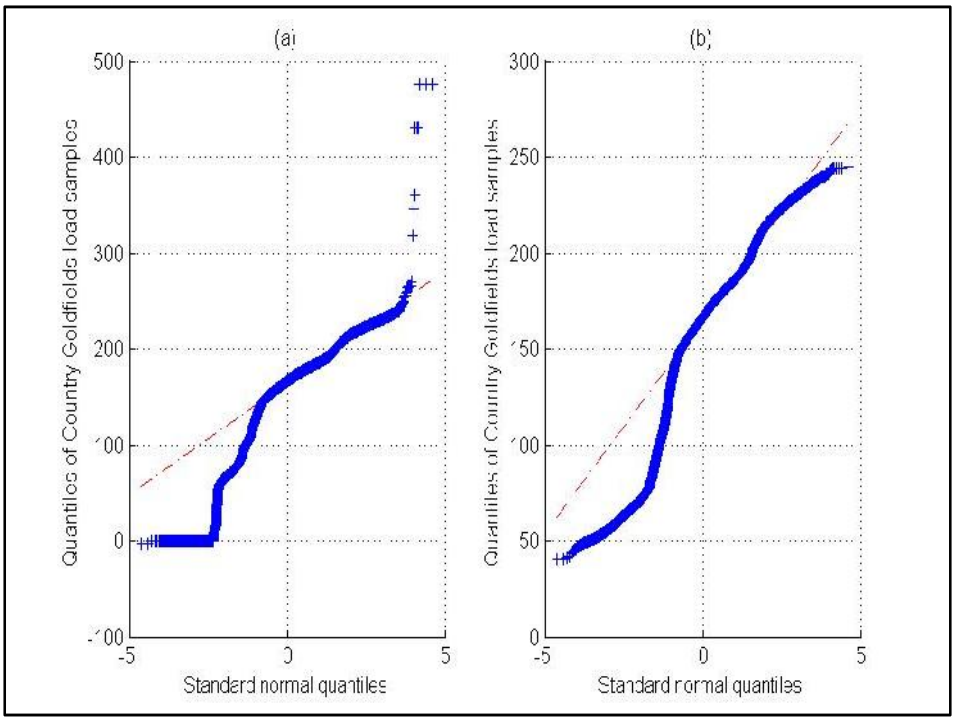


Figure 46: (a) Quantiles of Country Goldfields raw load versus the quantiles of standard normal; (b) Quantiles of Country Goldfields load with outliers and missing data points removed versus the quantiles of standard normal.

5.5 RESIDENTIAL, INDUSTRIAL, AND COMMERCIAL LOADS' BEHAVIOUR

The behaviours of residential, industrial, and commercial loads are different. Although in the practical case the load can be a combination of all the three types, it is important to study the properties of each separately. This section distinguishes these loads from each other. Finally a criterion will be proposed to recognise the dominancy of any of the mentioned types in a load data set. This criterion can be used in places where the dominant type of load is not known.

5.5.1 TEMPERATURE SENSITIVITY

Figure 47 to Figure 54 showeach region's load versus temperature for 15 years of observation. The red line in each graph roughly shows the regression between load and temperature during the hot and cold seasons[29]. In most cases, the slope of the line shows positive regression for the hot season and negative regression for the cold season. The difference is in the slope. The greatest the slope the more the temperature sensitivity.

Figure 47shows the scatter plot of Metro North load versus temperature. Positive regression for the hot season and negative regression for the cold season are very clear in this figure. As such, this load is highly temperature sensitive in all the seasons. The reason for this type of graph is

[29]The linear regression is used to visually inspect the temperature sensitivity of the load data. Obviously the relation between temperature and load data is not a linear one.

that Metro North is mainly a residential region, and people use electricity for cooling and heating purposes.

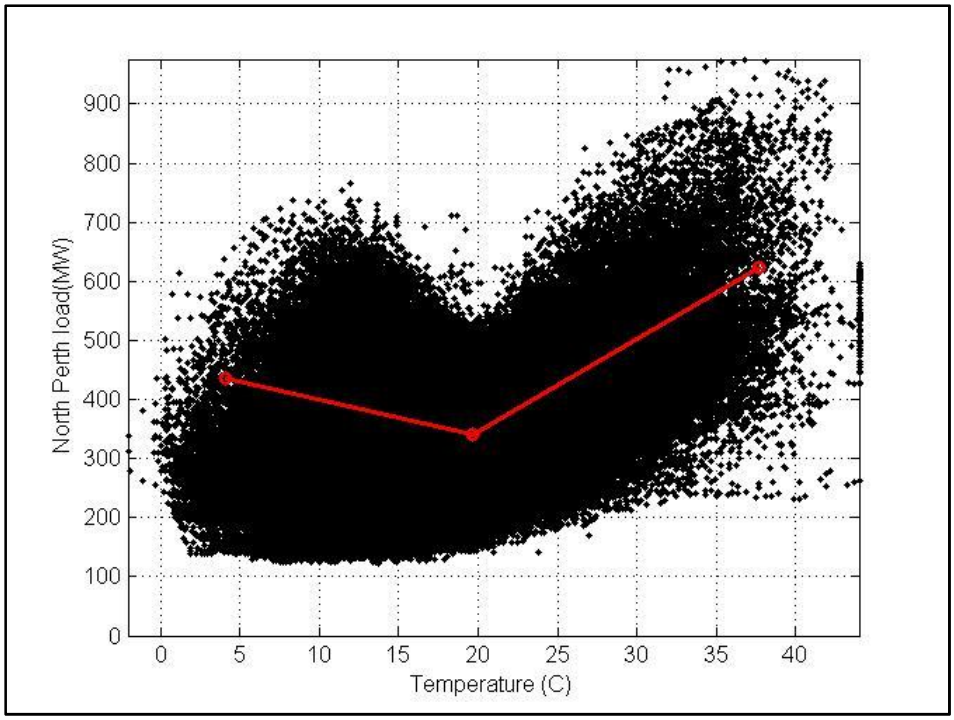


Figure 47: Half hourly load data (MW) versus temperature (degrees Celsius) of Metro North metropolitan area from 15 years of observation.

Figure 48 shows the scatter plot of Metro East load versus temperature. Similar to Figure 47, positive regression for the hot season and negative regression for the cold season are very clear in this figure. As such, this load is highly temperature sensitive in all the seasons. The reason for this type of graph is that Metro East is mainly a residential region, and people use electricity for cooling and heating purposes.

Figure 49shows the scatter plot of Metro South load versus temperature. Positive regression for the hot season and negative regression for the cold season are observable, although the slope of the line is less in the cold season. The reason for this type of graph is that Metro South is a residential and industrial region, but still mainly residential.

Figure 50 shows the scatter plot of the CBD load versus temperature. The line is similar to Figure 47. Although there is no industrial load in this region, there is a large commercial load. There are two main differences between commercial loads and residential loads. Commercial loads drop dramatically after business hours, and usually have lower heating demand and greater cooling demand. This while the left portion of the line that deals with heating is pretty flat in commercial regions. The main reason for this is the heating load, which is generated by electronic devices inside commercial buildings which needs to be removed by using

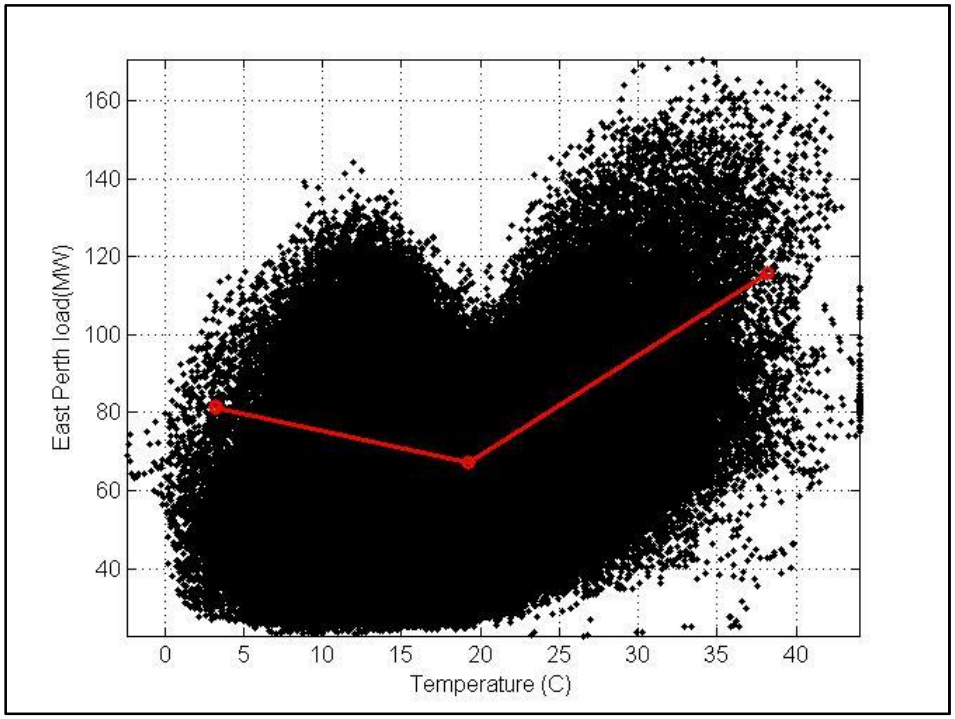


Figure 48: Half hourly load data (MW) versus temperature (degrees Celsius) of Metro East area from 15 years of observation.

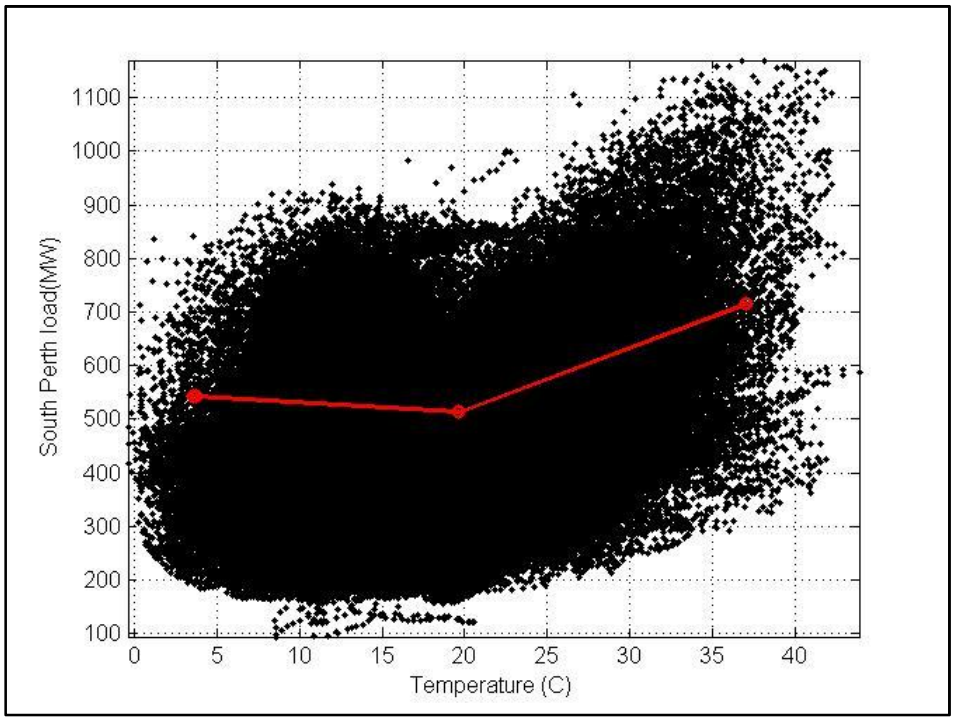


Figure 49: Half hourly load data (MW) versus temperature (degrees Celsius) of Metro South area from 15 years of observation.

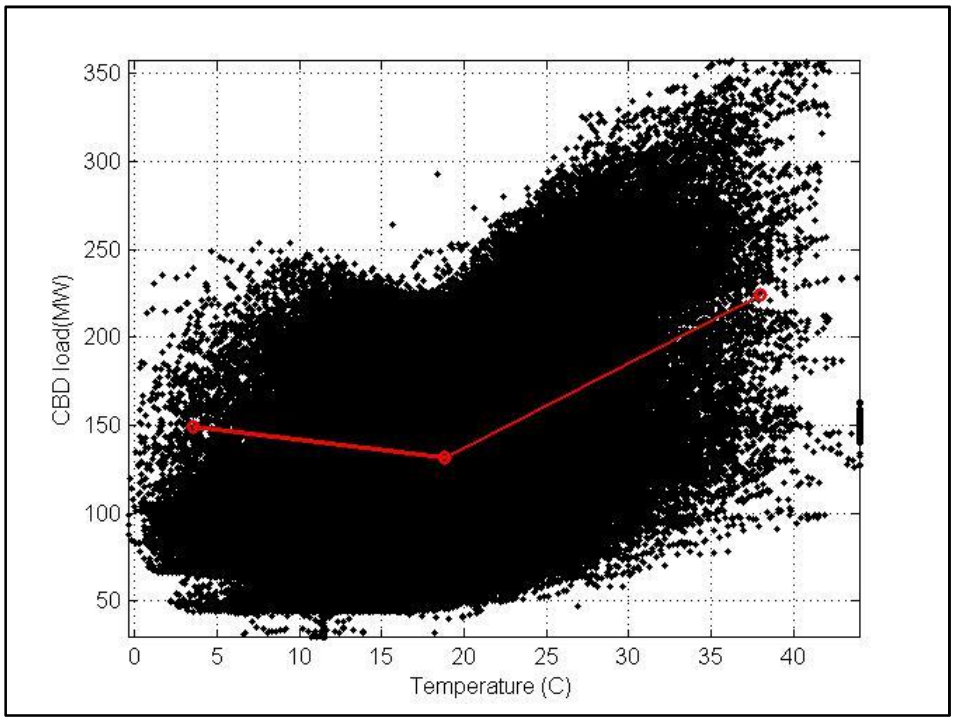


Figure 50: Half hourly load data (MW) versus temperature (degrees Celsius) of the CBD area from 15 years of observation.

air-conditioners. This increases the slope of the line in the hot season and decreases the slope in the cold season.

Figure 51shows the scatter plot of Country North load versus temperature. In this graph, the line is flat in the cold season and has a slight slope in the hot season. This indicates that there is a little temperature sensitivity in the hot season and almost no temperature sensitivity in the cold season. This region has both residential and industrial loads. As it is mostly dominated by industrial load, temperature sensitivity caused by residential component cannot be seen here. Figure 52 shows the scatter plot of Country East load versus temperature. This is similar to Figure 49, although there is a more visible slope in the cold season. This region has a greater residential load.

Figure 53 and Figure 54 respectively show the scatter plots of Country South and Country Goldfields load versus temperature. The lines are almost flat in these graphs, which indicate negligible temperature sensitivity. Irrespective of the outside temperature, electricity load varies based on the factory demand. Both of these regions are dominated by industrial loads.

It can be concluded that more residential load in a region will introduce more temperature sensitivity and that more industrial load will reduce it. The behaviour of commercial loads is similar to residential loads in the hot season and similar to industrial loads in the cold season.

Notably, these conclusions are only valid for regions where electricity is used for both cooling and heating purposes. If electricity is not used for heating purposes in the cold season or for

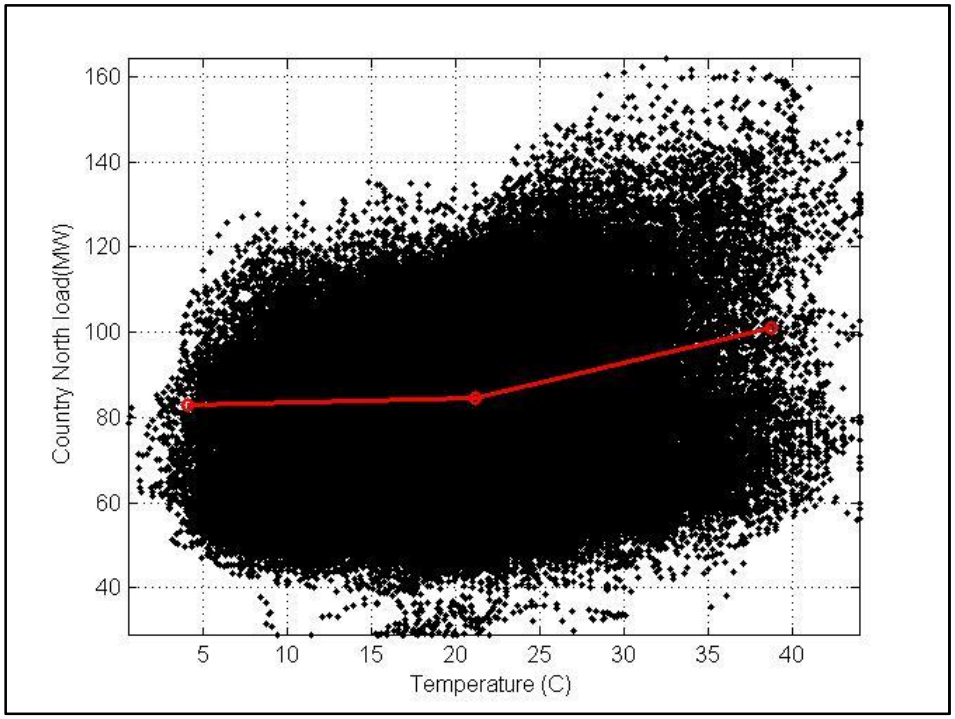


Figure 51: Half hourly load data (MW) versus temperature (degrees Celsius) of Country North area from 15 years of observation.

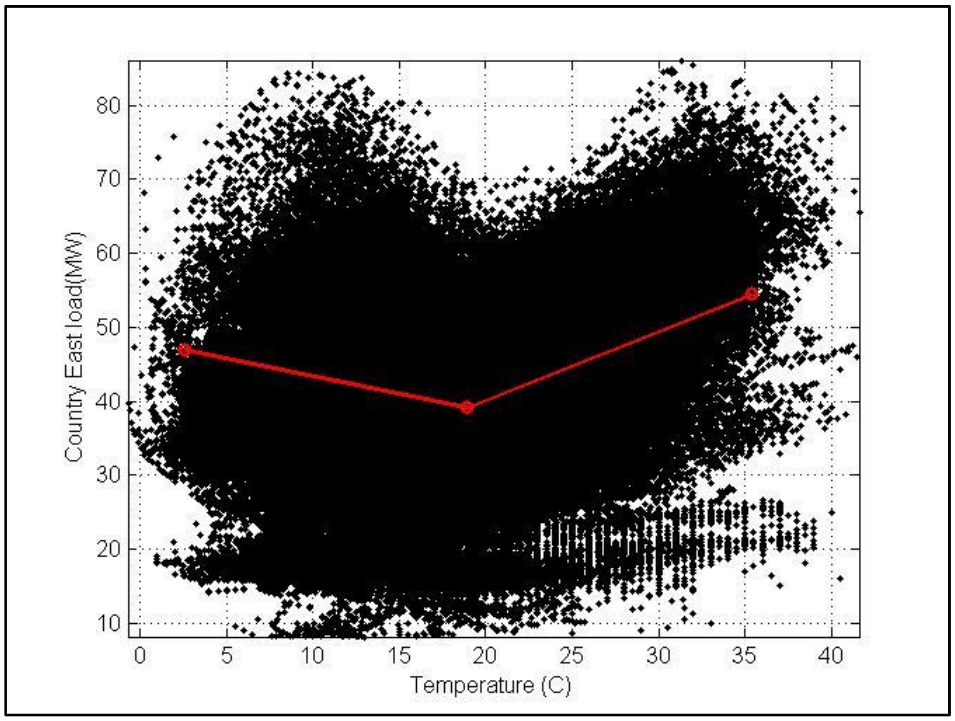


Figure 52: Half hourly load data (MW) versus temperature (degrees Celsius) of Country East area from 15 years of observation.

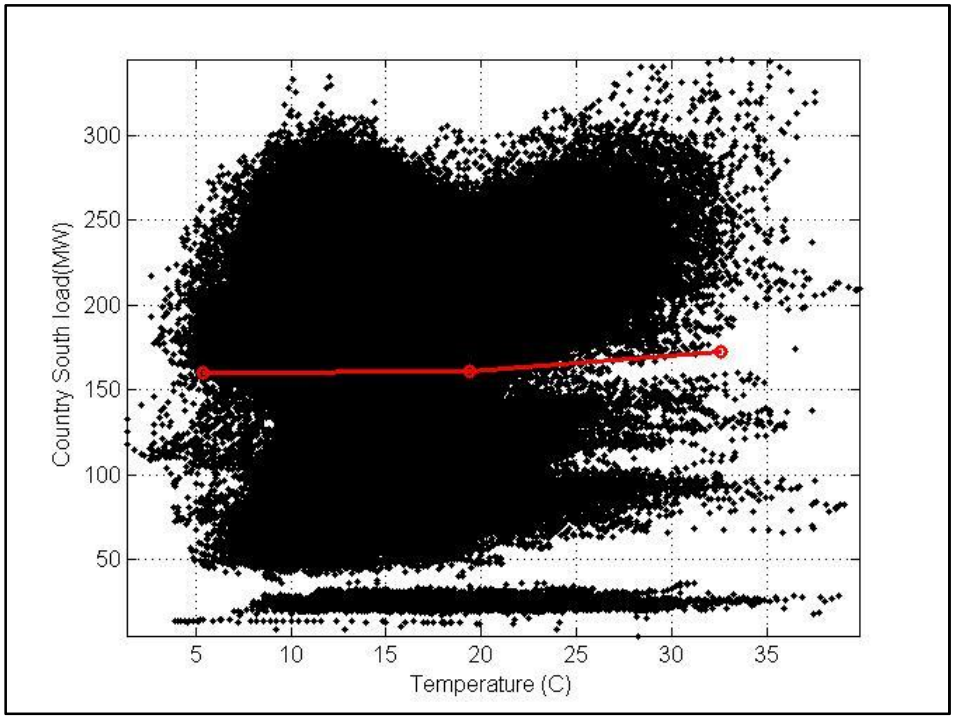


Figure 53: Half hourly load data (MW) versus temperature (degrees Celsius) of Country South area from 15 years of observation.

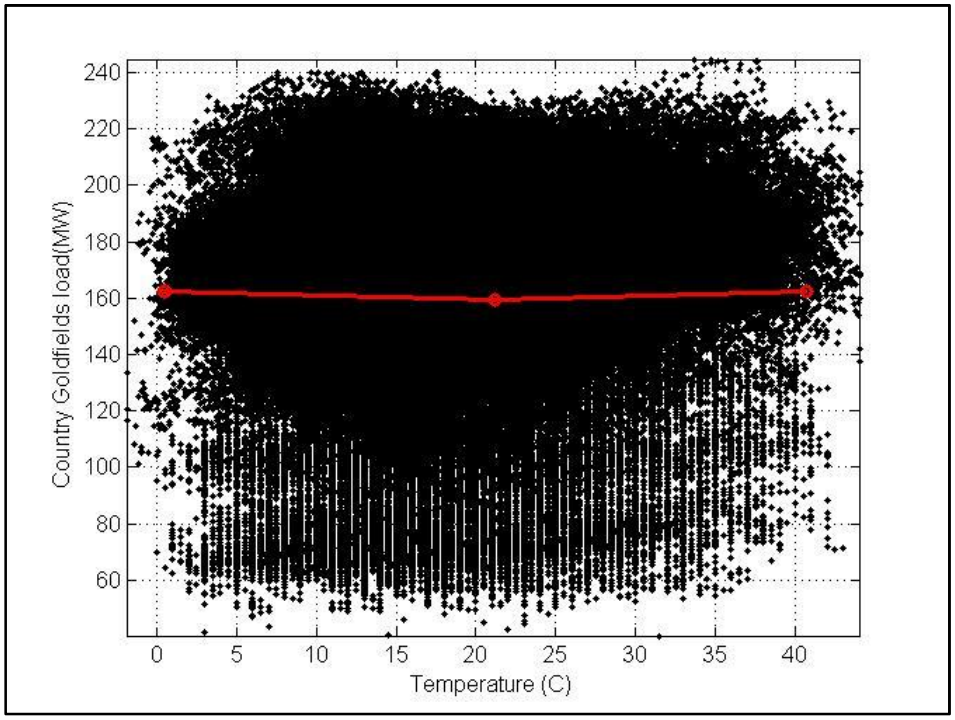


Figure 54: Half hourly load data (MW) versus temperature (degrees Celsius) of Country Goldfields area from 15 years of observation.

cooling purposes in the hot season then the temperature-load regression for the season will be zero (flat line). The latter case is very rare, although an example is the use of absorption chillers in the hot season.

5.5.2 DISTRIBUTION

This section will study the behaviours of different loads based on their distributions. As discussed in 5.3.2 Q-Q plots can be used to determine whether two data sets come from populations with a common distribution. This section will investigate this more and will come up with an invaluable criterion which can be used on load data to determine which of the load components (commercial, residential or industrial) are more dominant in the data set[30].

To capture useful information from the available load data, Q-Q plots of them versus eight different distributions have been generated. The selected distributions are [122]: standard normal, Gamma, lognormal, Weibull, generalised extreme value (GEV), generalised pareto (GP), Poisson, and Rayleigh (R). The number of available probability distributions in the literature is countless. The above eight distributions are selected after a few tests on a bigger list of distributions. Each of the selected distributions at least fit pretty well with one of our datasets. For example Figure 55 shows a reasonable good fit between Metro North load and a few of selected distributions such as standard normal, Gamma, Lognormal, Weibull, and GEV. But not a good fit with GP, Poisson and Rayleigh. Good fit with GP can be seen in Figure 58 with quantiles of CBD load and so on.

Figure 55 to Figure 62 show the Q-Q plots for each region's load versus the mentioned distributions for a representative year.

[30]To date, such behavioural analysis of load data has not been addressed by any other sources available in the literature.

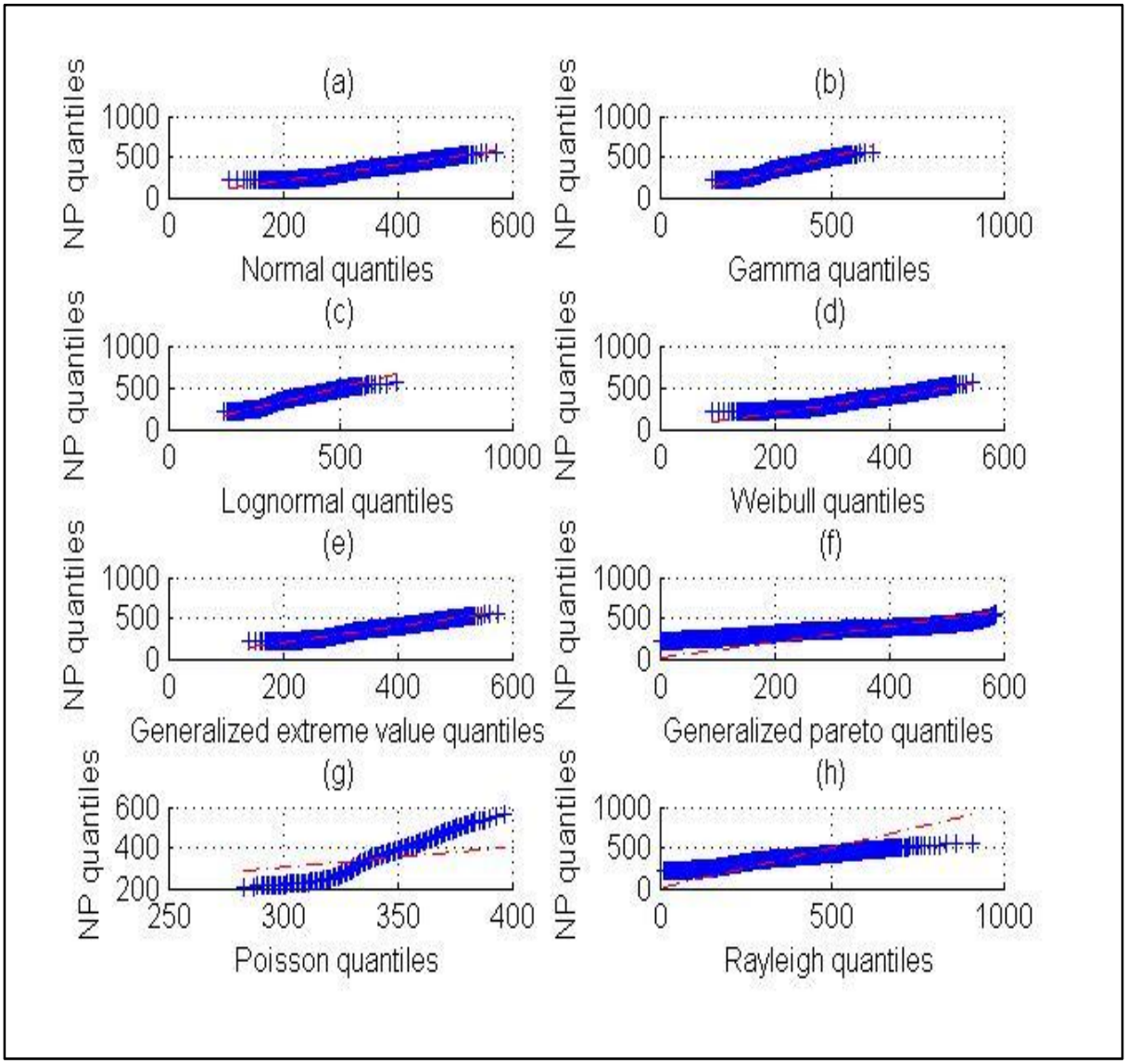


Figure 55: Plot of the quantiles of Metro North load versus various distributions. (a) Q-Q plot of Metro North load versus normal distribution; (b) Q-Q plot of Metro North load versus gamma distribution; (c) Q-Q plot of Metro North load versus lognormal distribution; (d) Q-Q plot of Metro North load versus Weibull distribution; (e) Q-Q plot of Metro North load versus generalised extreme value distribution; (f) Q-Q plot of Metro North load versus generalised Pareto distribution; (g) Q-Q plot of Metro North load versus Poisson distribution; (h) Q-Q plot of Metro North load versus Rayleigh distribution.

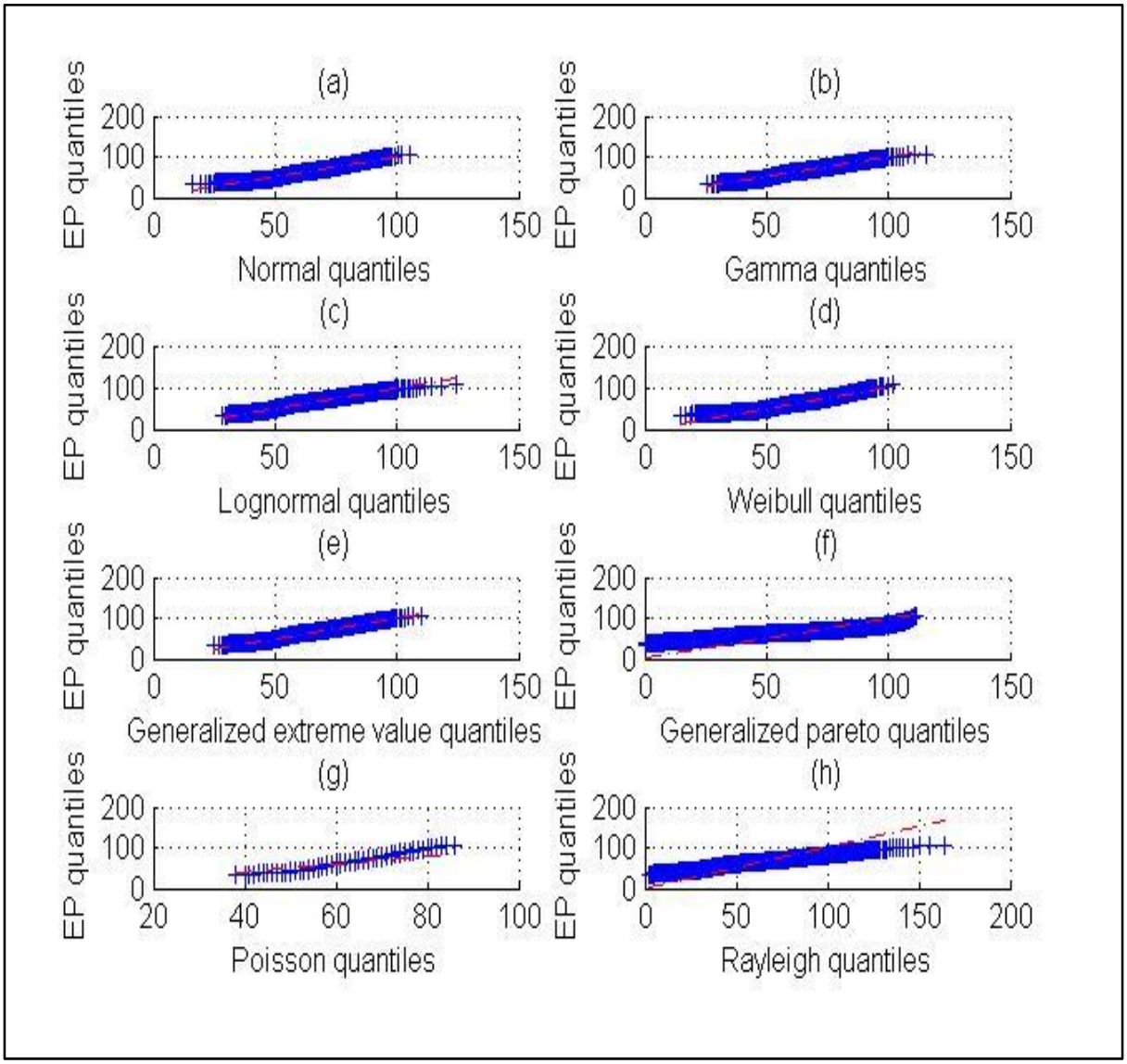


Figure 56: Plot of the quantiles of Metro East load versus various distributions. (a) Q-Q plot of Metro East load versus normal distribution; (b) Q-Q plot of Metro East load versus gamma distribution; (c) Q-Q plot of Metro East load versus lognormal distribution; (d) Q-Q plot of Metro East load versus Weibull distribution; (e) Q-Q plot of Metro East load versus generalised extreme value distribution; (f) Q-Q plot of Metro East load versus generalised Pareto distribution; (g) Q-Q plot of Metro East load versus Poisson distribution; (h) Q-Q plot of Metro East load versus Rayleigh distribution.

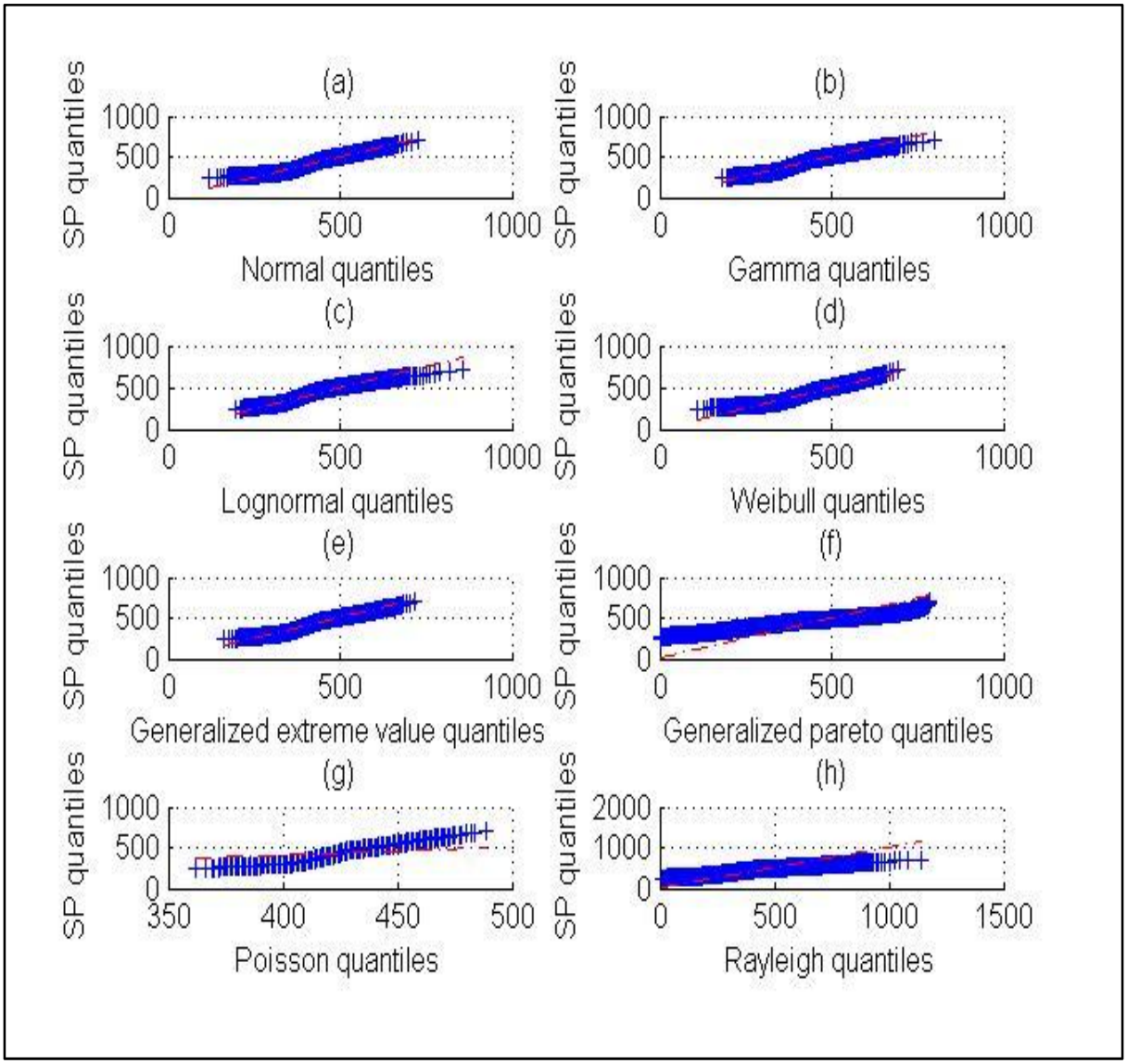


Figure 57: Plot of the quantiles of Metro South load versus various distributions. (a) Q-Q plot of Metro South load versus normal distribution; (b) Q-Q plot of Metro South load versus gamma distribution; (c) Q-Q plot of Metro South load versus lognormal distribution; (d) Q-Q plot of Metro South load versus Weibull distribution; (e) Q-Q plot of Metro South load versus generalised extreme value distribution; (f) Q-Q plot of Metro South load versus generalised Pareto distribution; (g) Q-Q plot of Metro South load versus Poisson distribution; (h) Q-Q plot of Metro South load versus Rayleigh distribution.

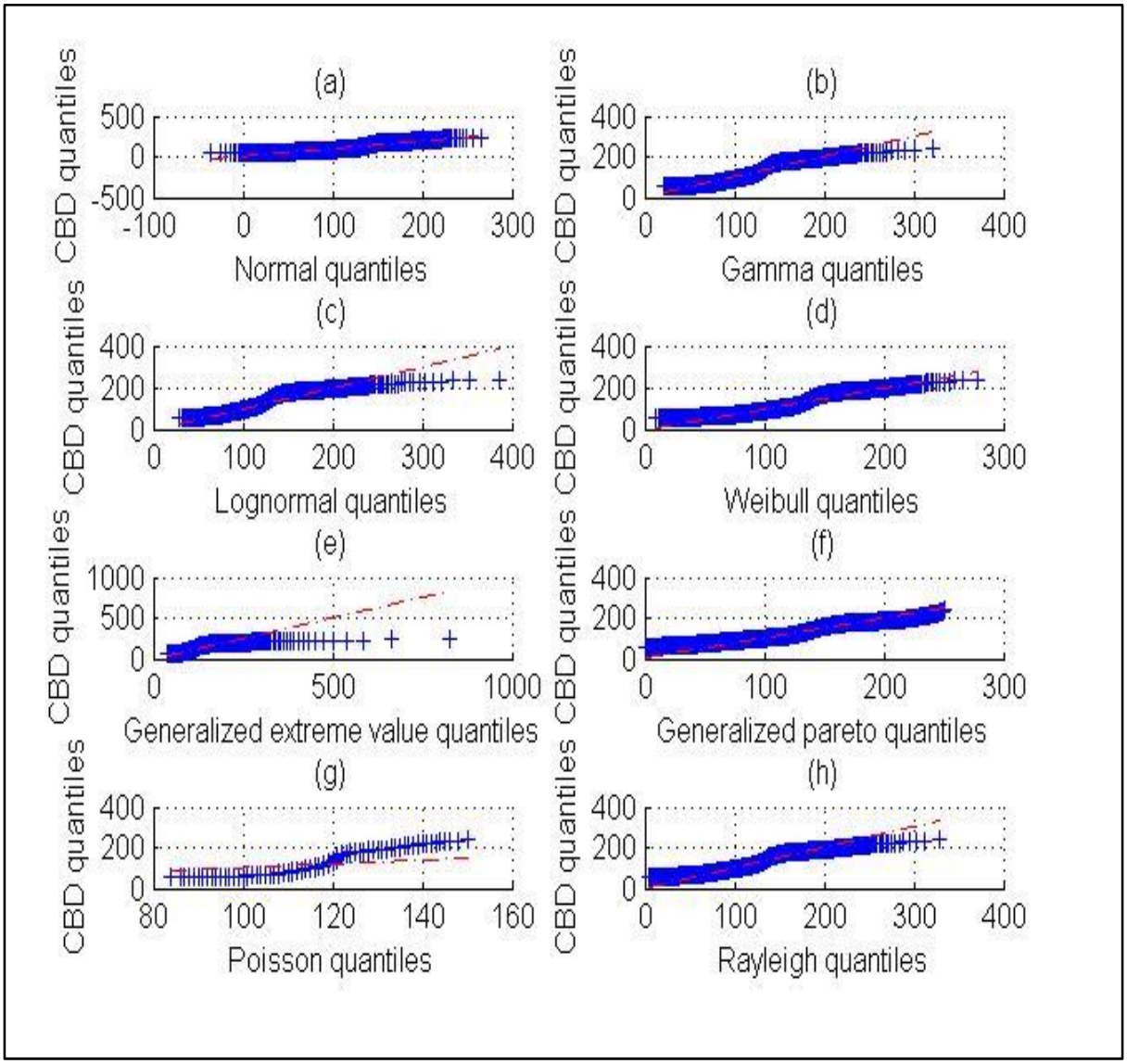


Figure 58: Plot of the quantiles of the CBD load versus various distributions. (a) Q-Q plot of CBD load versus normal distribution; (b) Q-Q plot of CBD load versus gamma distribution; (c) Q-Q plot of CBD load versus lognormal distribution; (d) Q-Q plot of CBD load versus Weibull distribution; (e) Q-Q plot of CBD load versus generalised extreme value distribution; (f) Q-Q plot of CBD load versus generalised Pareto distribution; (g) Q-Q plot of CBD load versus Poisson distribution; (h) Q-Q plot of CBD load versus Rayleigh distribution.

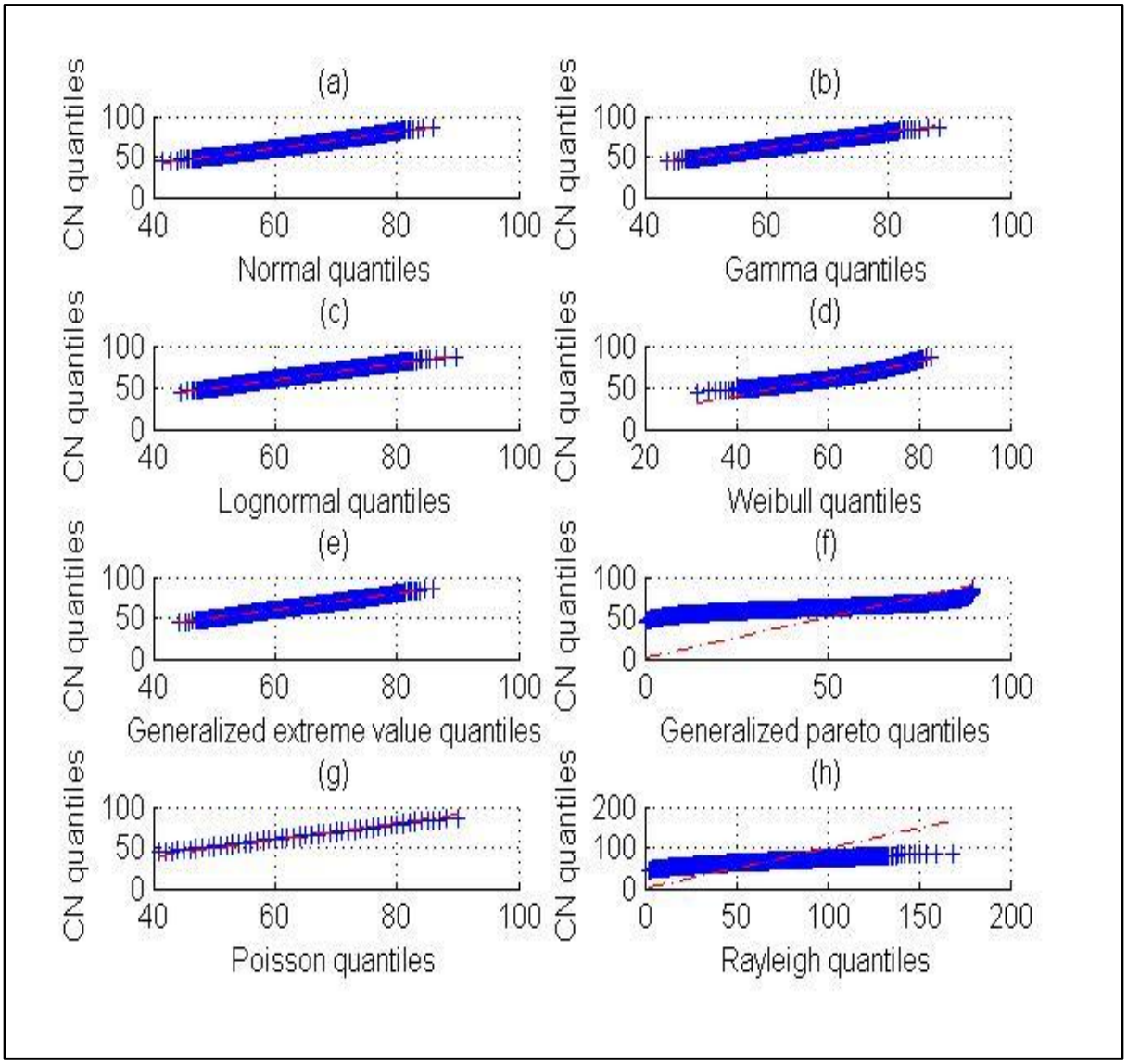


Figure 59: Plot of the quantiles of Country north region load versus various distributions. (a) Q-Q plot of Country north region load versus normal distribution; (b) Q-Q plot of Country north region load versus gamma distribution; (c) Q-Q plot of Country north region load versus lognormal distribution; (d) Q-Q plot of Country north region load versus Weibull distribution; (e) Q-Q plot of Country north region load versus generalised extreme value distribution; (f) Q-Q plot of Country north region load versus generalised Pareto distribution; (g) Q-Q plot of Country north region load versus Poisson distribution; (h) Q-Q plot of Country north region load versus Rayleigh distribution.

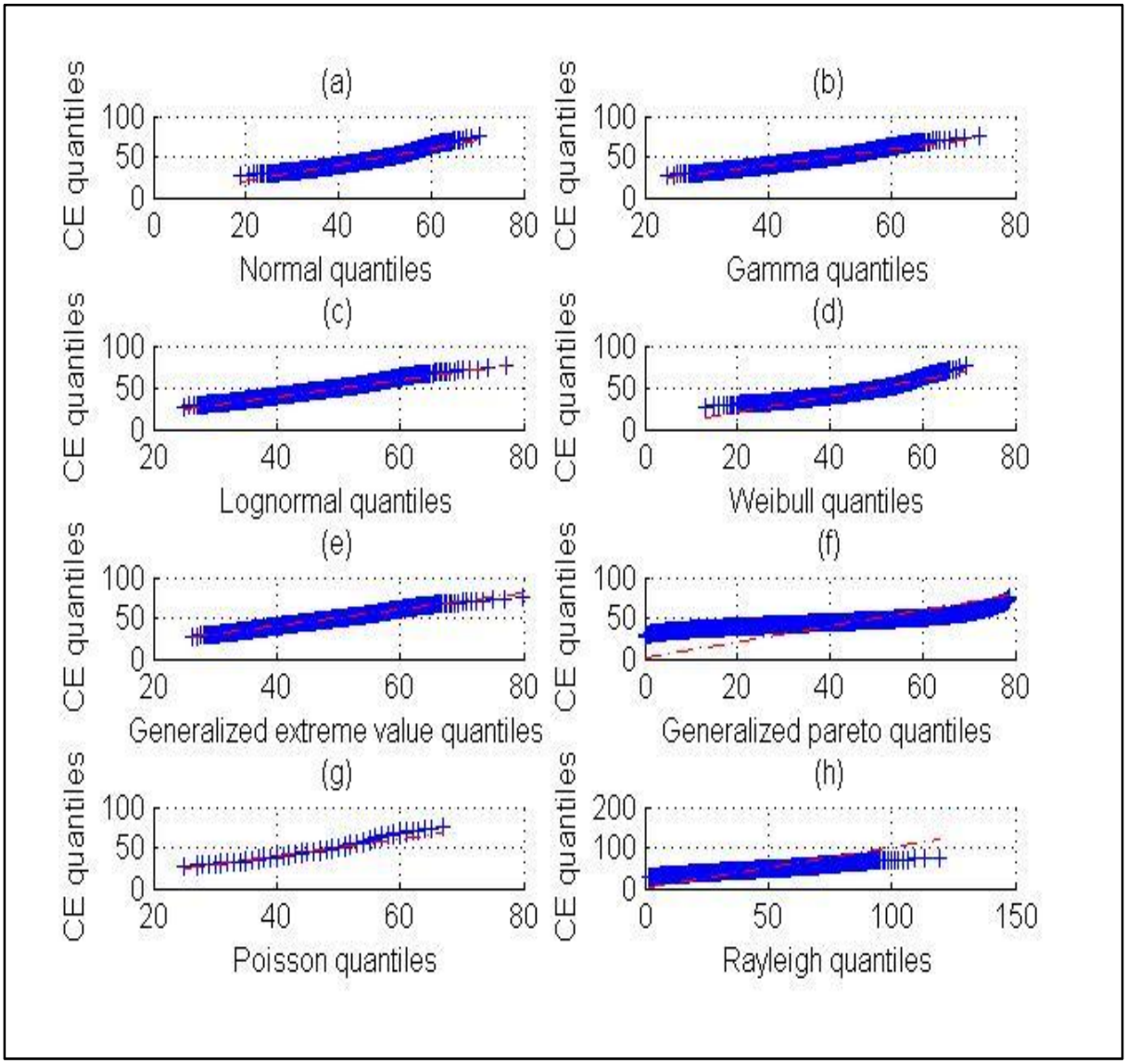


Figure 60: Plot of the quantiles of Country east region load versus various distributions. (a) Q-Q plot of Country east region load versus normal distribution; (b) Q-Q plot of Country east region load versus gamma distribution; (c) Q-Q plot of Country east region load versus lognormal distribution; (d) Q-Q plot of Country east region load versus Weibull distribution; (e) Q-Q plot of Country east region load versus generalised extreme value distribution; (f) Q-Q plot of Country east region load versus generalised Pareto distribution; (g) Q-Q plot of Country east region load versus Poisson distribution; (h) Q-Q plot of Country east region load versus Rayleigh distribution.

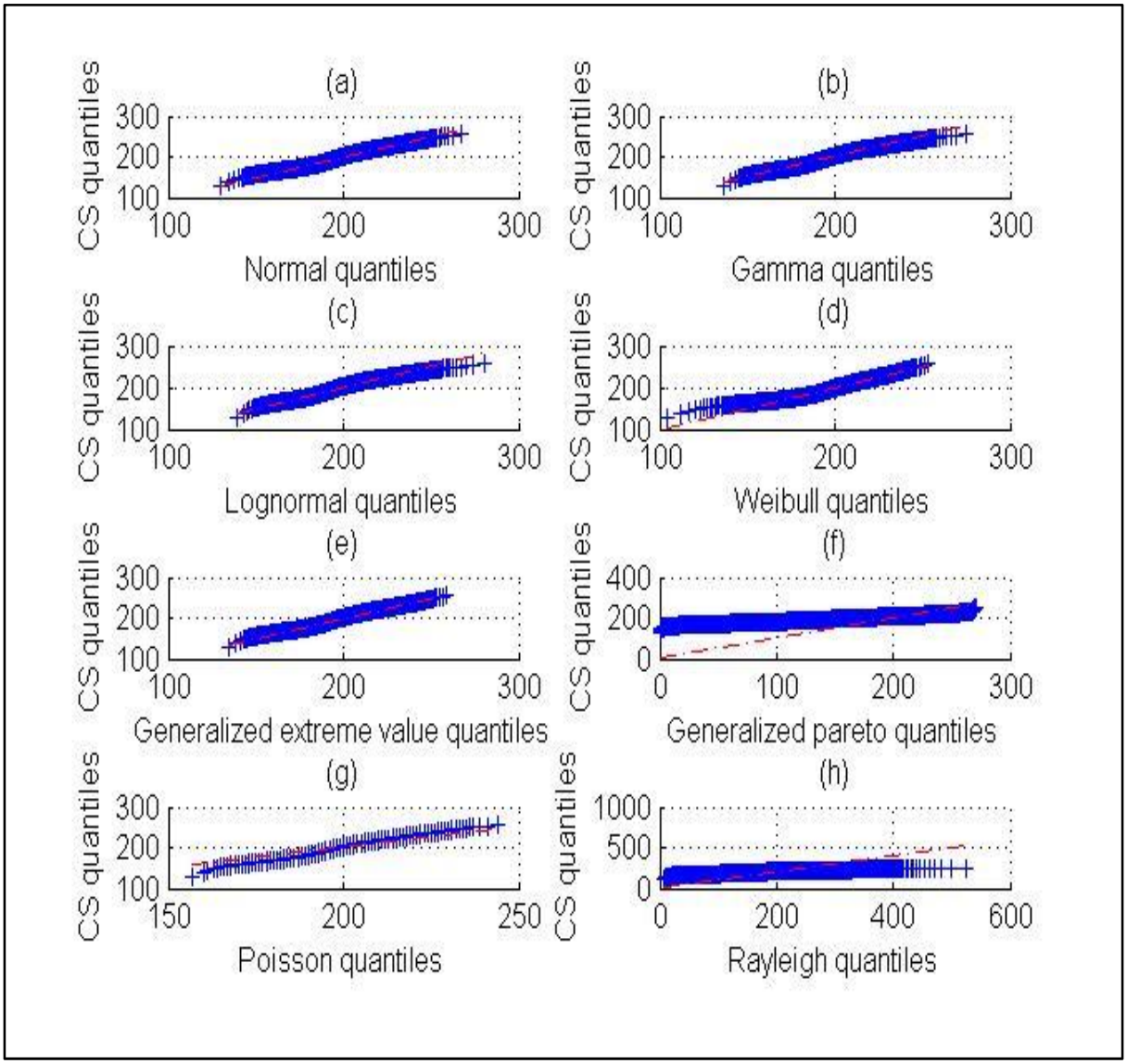


Figure 61: Plot of the quantiles of Country South region load versus various distributions. (a) Q-Q plot of Country South region load versus normal distribution; (b) Q-Q plot of Country South region load versus gamma distribution; (c) Q-Q plot of Country South region load versus lognormal distribution; (d) Q-Q plot of Country South region load versus Weibull distribution; (e) Q-Q plot of Country South region load versus generalised extreme value distribution; (f) Q-Q plot of Country South region load versus generalised Pareto distribution; (g) Q-Q plot of Country South region load versus Poisson distribution; (h) Q-Q plot of Country South region load versus Rayleigh distribution.

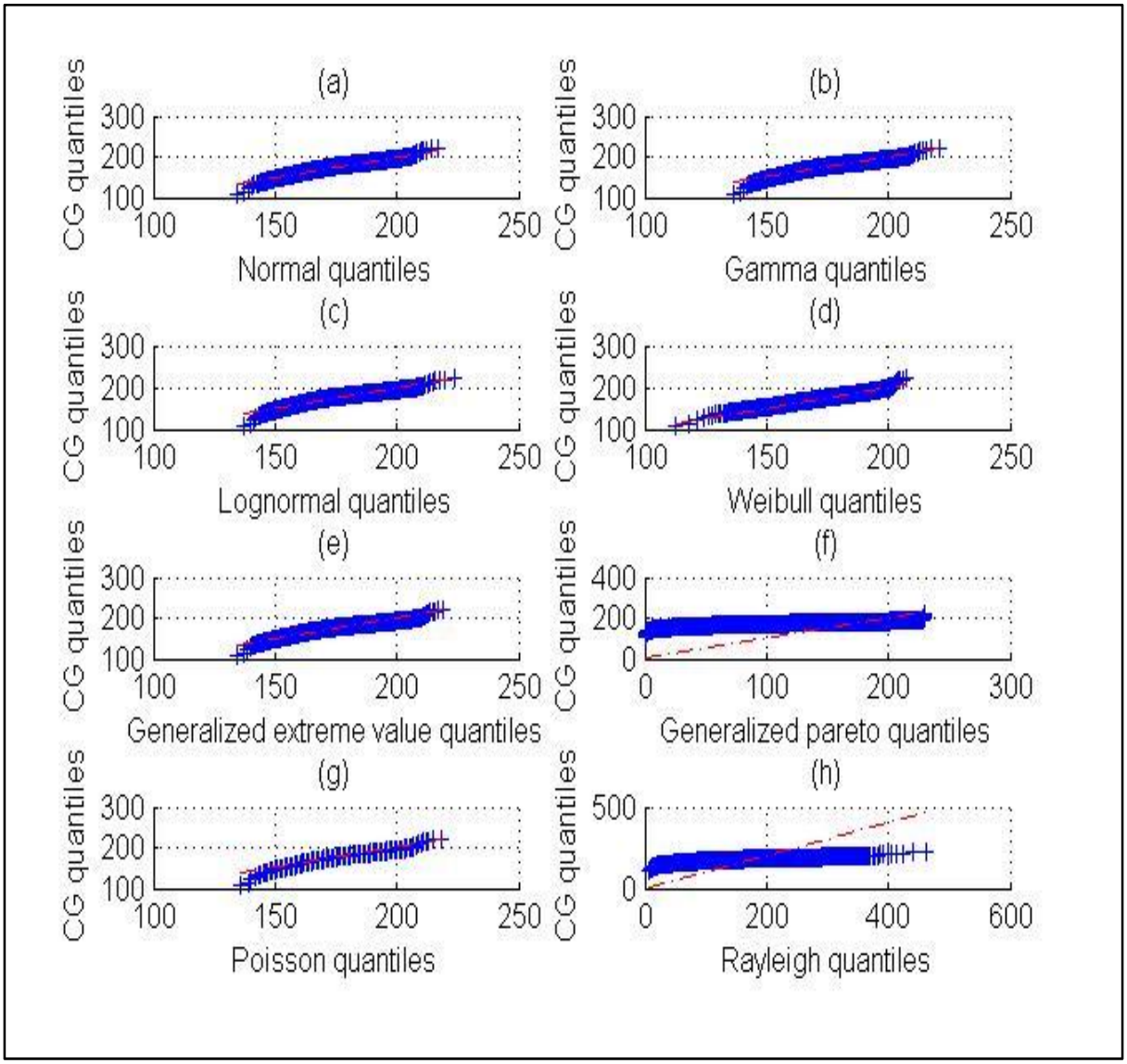


Figure 62: Plot of the quantiles of Country Goldfields region load versus various distributions. (a) Q-Q plot of Country Goldfields region load versus normal distribution; (b) Q-Q plot of Country Goldfields region load versus gamma distribution; (c) Q-Q plot of Country Goldfields region load versus lognormal distribution; (d) Q-Q plot of Country Goldfields region load versus Weibull distribution; (e) Q-Q plot of Country Goldfields region load versus generalised extreme value distribution; (f) Q-Q plot of Country Goldfields region load versus generalised Pareto distribution; (g) Q-Q plot of Country Goldfields region load versus Poisson distribution; (h) Q-Q plot of Country Goldfields region load versus Rayleigh distribution.

Although some of the selected distributions show pretty good fits with the load data available (see Figure 55 to Figure 62) , the goal here is to be able to distinguish between the available load data. For example standard normal quantiles and Gamma quantiles show good fits in many regions and for the same reason they cannot be used to help us distinguish between the load data available. Among all the distributions, three carry vital information that help to distinguish

the dominant component of the electricity load. Those three are Rayleigh (R), Generalised Pareto (GP), and Generalised Extreme Value (GEV).

To carry on with the test, three regions of East Perth, CBD and Country Goldfields are selected as samples of pure residential, commercial and industrial loads respectively. Figure 63 and Figure 64 illustrate the Q-Q plots of all three types of load versus R and GP distribution. In both figures, the best fit is for commercial load. The residential load is not completely fitted, but is a fairly good fit compared to the industrial load, which shows a totally different distribution.

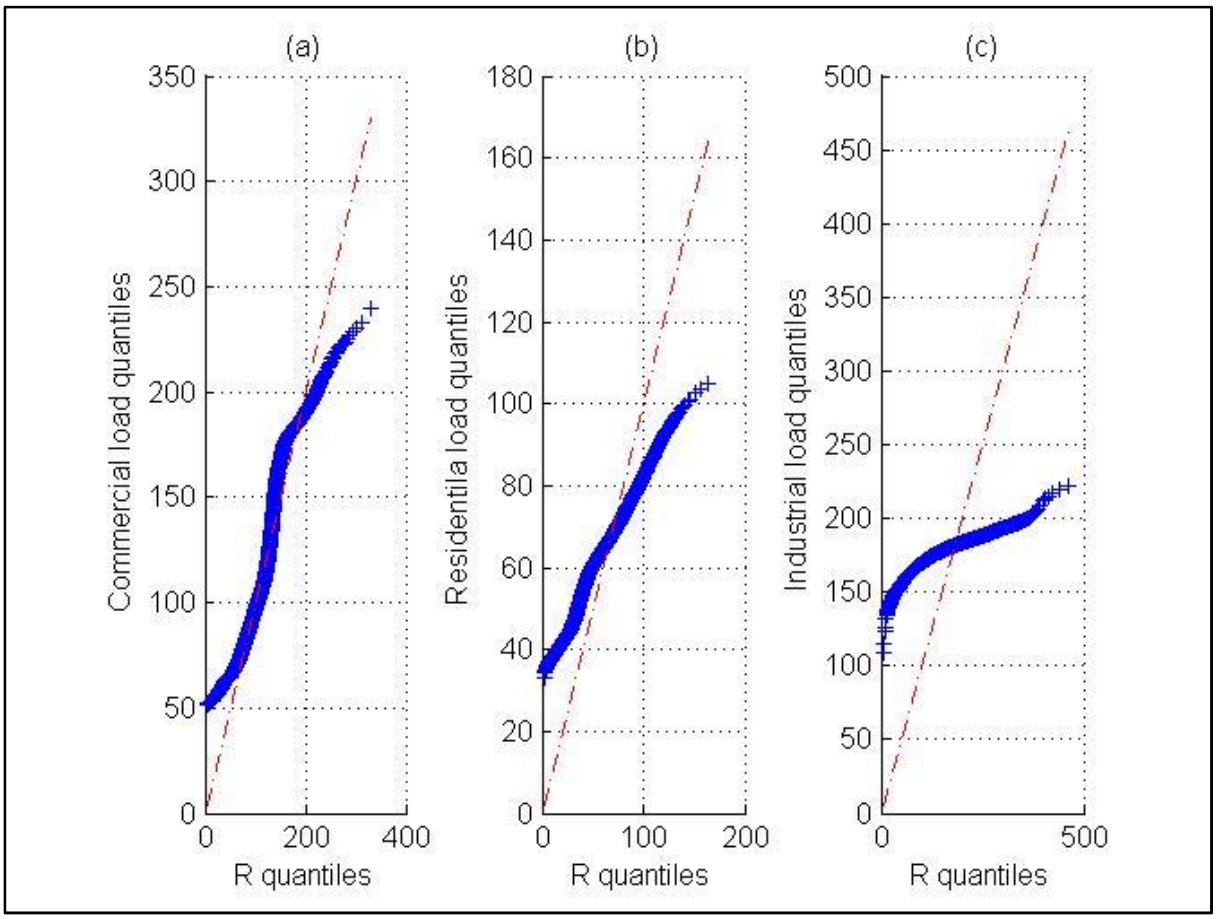


Figure 63: (a) Q-Q plot of commercial load (CBD region) versus Rayleigh distribution; (b) Q-Q plot of residential load (East Perth region) versus Rayleigh distribution; (c) Q-Q plot of industrial load (Country Goldfilelds region) versus Rayleigh distribution.

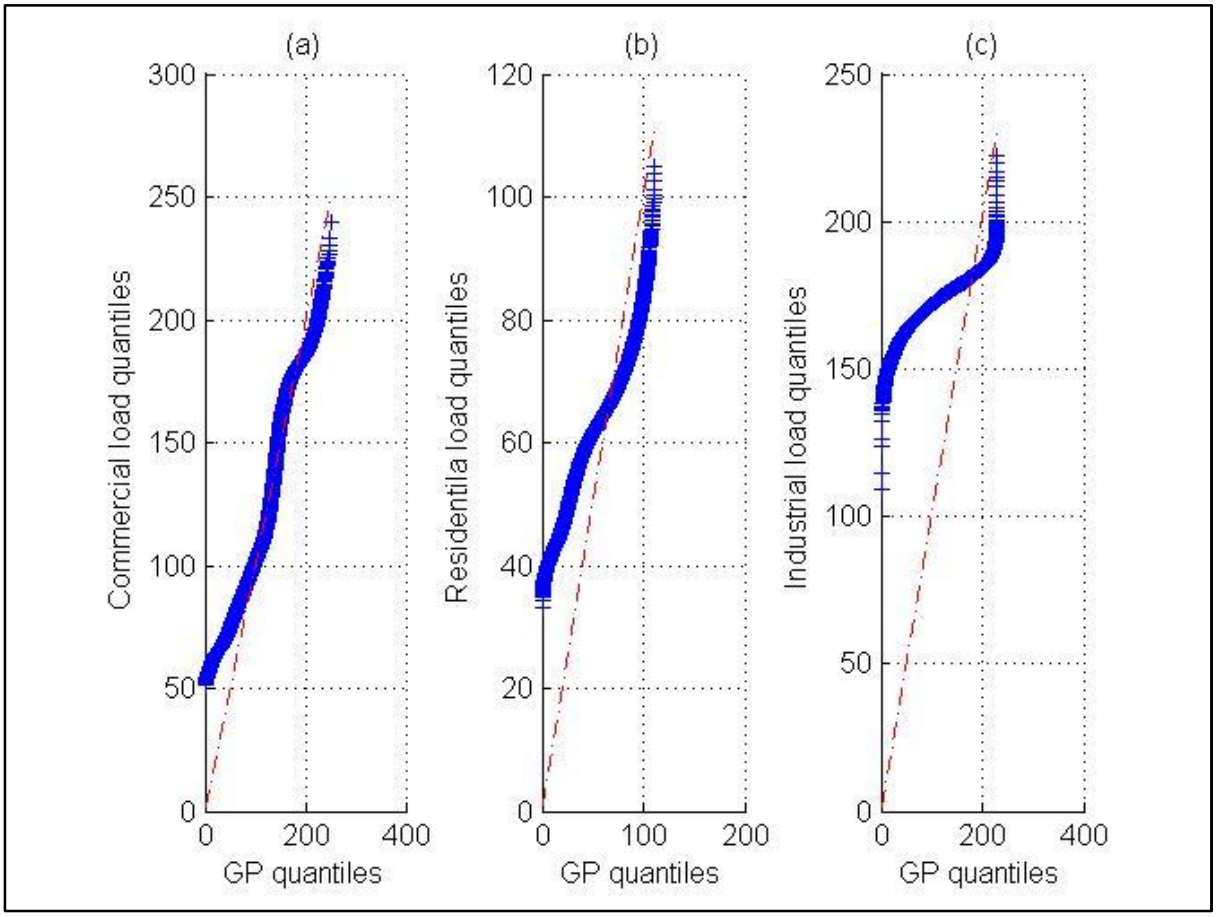


Figure 64: (a) Q-Q plot of commercial load (CBD region) versus generalised Pareto distribution; (b) Q-Q plot of residential load (East Perth region) versus generalised Pareto distribution; (c) Q-Q plot of industrial load (Country Goldfield region) versus generalised Pareto distribution.

The GEV distribution is shown in Figure 65. Unlike the other two distributions, the fit is very good for the industrial and residential loads. The commercial load cannot be fitted by this type of distribution.

Based on these observations, a load type determination criterion can be developed. The user of this criterion may plot the load versus Rayleigh(R), Generalised Pareto(GP), and Generalised Extreme Value(GEV) distributions, and compare the output with the general rule presented in Table 5 to find the dominant component of the load. Notably, this criterion can be only be applied to areas where both electrical heating and cooling are being used by the customers.

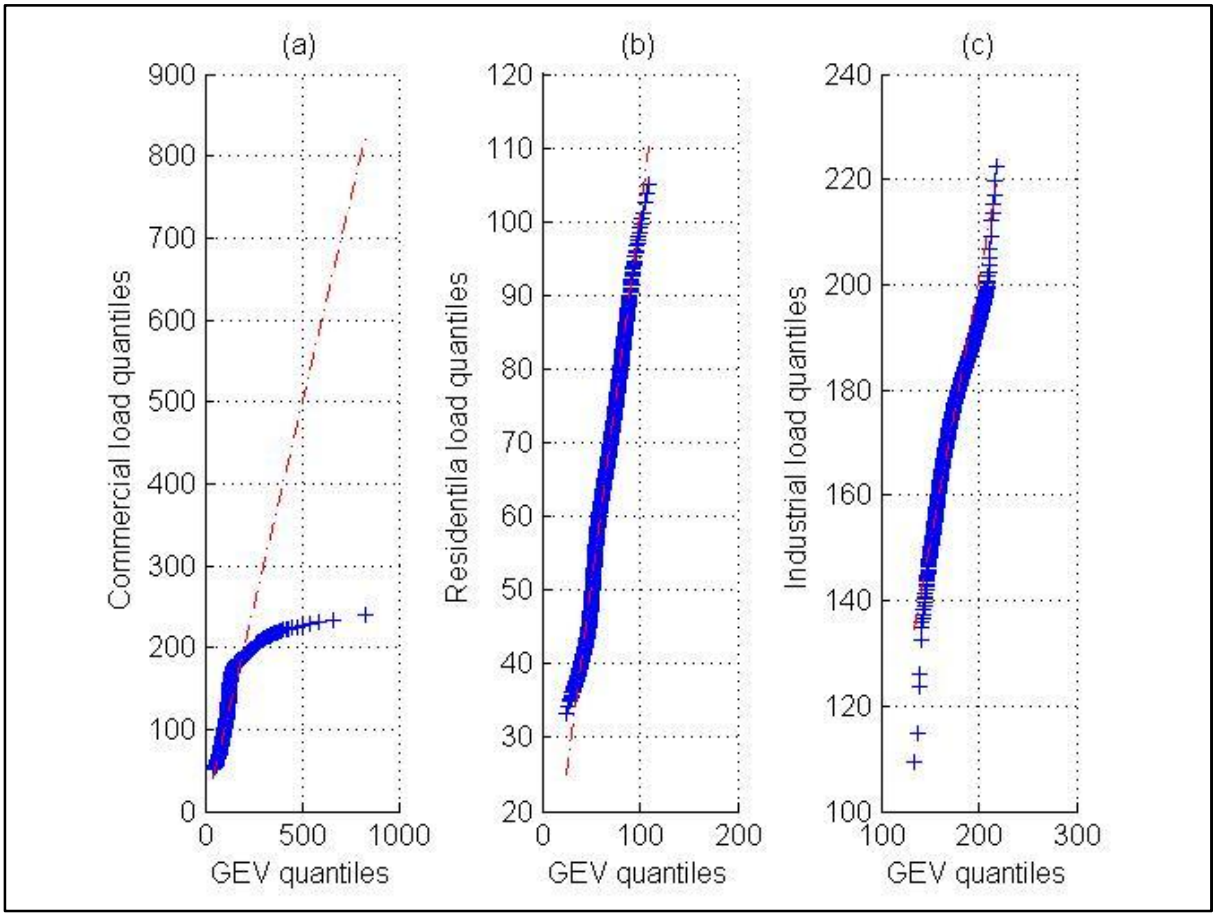


Figure 65: (a) Q-Q plot of commercial load (CBD region) versus generalised extreme value distribution; (b) Q-Q plot of residential load (East Perth region) versus generalised extreme value distribution; (c) Q-Q plot of industrial load (Country Goldfields region) versus generalised extreme value distribution.

Table 5: Load type determination criterion.

Load type	Commercial	Residential	Industrial
Rayleigh	Good	Fairly good	Bad
Generalised Pareto	Good	Fairly good	Bad
Generalised Extreme Value	Bad	Good	Good

5.6 FORECASTED RESULTS

After pre-processing and clustering the data, 13 sets of input variables were defined based on the available temporal and weather data as the feed for training models. Preparing a proper set of input variables is a significant step in any training procedure and can strongly affect the accuracy of the method. Proper input data includes variables that correlate with the output data or variables that help to classify the other input variables. The input variables are year, month, day of the week, hour of the day, temperature, relative humidity, previous-day same-hour

demand, previous-week same-hour demand, holidays,[31] average past-24-hour demand, average past-seven-day demand, summer temperature to distinguish the hot-day temperature, and winter temperature to distinguish the cold-day temperature.

The set of input variables to training models consists of 13 column vectors of input variables with a total of 275,513[32] observations in each vector and one column vector of the same size for target variables.

Two different nonlinear training methods, namely, artificial neural networks (ANN) and decision trees learning, were used in this study. Because of the poor performance of ARX on the available data set, as investigated in 4.6, it is excluded from the training methods used. Bagging decision trees was selected for the construction of ensembles due to its performance under classification noise[111]. To construct more than one tree in regression tree method, techniques called ensemble methods can be used. The ensemble method can be bagging [109], random forest, boosted trees [110] and rotation forest. Due to its performance under classification noise, bagging has been selected for the construction of regression trees for ensembles as reported in [111].Input and target variables for the training period were used to find the optimum configuration for ANN and bagging decision trees.

The models were trained with 14 years of data from April 1995 to December 2009. For all cases, the testing period is 2010. Upon completion of the training procedures, the trained models can be used for future simulations. Using the single-day forecast, and the reconstructed temperature and humidity signals (see 4.3.3 Clustering and signal reconstruction) as new inputs for the trained models, the forecasting horizon can be stretched from one day to seven days.

The architecture of the neural network used is feed forward back-propagation with 40 hidden layers and one output layer. The number of layers and type of feed forward network is determined after running few tests on load data from 1995 to 2008. This configuration had the best regression for both training sets and validation sets. Training performance is set on minimising the mean absolute error.

Decision trees architecture used considers 40 number of regression tree with 30 as the minimum leaf size. Similar to neural network architecture, these numbers are found by trial and error on the data from 1995 to 2008. See codes for more details.

Given that different regions have different average load, mean absolute percentage error (MAPE) may not provide a good comparison. For a better comparison of the performance of the models out of sample data (during the test year), modified mean absolute percentage error (MMAPE) has been defined in equation (1).

Equation (1) basically multiplies the MAPE value by a coefficient, which is a function of average load in different regions. The resulting MMAPE is no longer affected by the average load of the

[31]A list of Western Australian public holidays has been used to generate the holidays input variable.

[32]Half an hourly samples from April 1995 to December 2009

region itself. In simple-terms MMAPE normalizes MAPE across all the regions and puts all the MAPEs on the same scale for comparison between them.

$$MMAPE_z = \frac{\frac{T}{N}\sum_{t=1}^{N} A_{zt}}{\sum_{i=1}^{T}\frac{1}{N}\sum_{t=1}^{N} A_{it}} \frac{1}{N}\sum_{t=1}^{N}\left|\frac{A_{zt}-F_{zt}}{A_{zt}}\right|$$

A : *Actual load*

F : *Forecasted load* (5-1)

N : *Number of samples*

T : *Number of all the regions except z*

For both models, the daily and weekly MMAPE were calculated for each month of the test period. The results are shown in Table 6 to Table 13.

All these tables consist of twelve rows and five columns. Each row is representative of a month of the year. The first two columns contain the daily MMAPE percentage of neural network and decision trees models, respectively. The next two represent the weekly MMAPE of those methods and the last column illustrates the average temperature of that specific month over the training period.

Table 6belongs to Metro North where consumers are a combination of residential and industrial ones. All of the daily MMAPE values are below 5%. For the months of Jan to March neural network shows a better result compared to decision trees. For the rest of the year decision trees accuracies are superior. In other words in hot months[33] of the year neural network performs better than decision trees.

[33]Hot weather is when an occupant of a building feels the need to turn on the air-conditioning unit. Human conditions of comfort depends on many different items and it is explained in a very good level of details by ASHRAE Standard 55–thermal environmental conditions for human occupancy[137]. In this study we assume any dry bulb temperature above 23°C to be hot.

Table 6: MMAPE of Metro North out of sample data (2010 as the test year).

	Daily MMAPE NN	Daily MMAPE DT	Weekly MMAPE NN	Weekly MMAPE DT	Average Temperature (°C)
Jan	3.7%	4.9%	4.5%	5.6%	25.4
Feb	3.3%	4.5%	4.2%	5.3%	24.4
Mar	4.0%	4.6%	4.9%	5.5%	23.2
Apr	2.7%	2.5%	3.5%	3.1%	18.8
May	3.5%	2.4%	4.1%	3.2%	14.7
Jun	3.4%	2.8%	4.1%	3.8%	11.8
Jul	3.2%	2.6%	4.3%	3.5%	11.3
Aug	2.7%	2.5%	3.3%	3.1%	12.3
Sep	2.8%	2.3%	3.7%	3.1%	14.8
Oct	2.4%	1.6%	3.2%	2.8%	17.5
Nov	3.9%	3.4%	5.2%	4.7%	21.9
Dec	4.8%	3.7%	5.5%	4.5%	22.5

Table 7 belongs to Metro East where the consumers of electricity are purely residential. All of the daily MMAPE values are below 3% which is much lower that previous table. For the months of Nov to Feb neural network shows a better result compared to decision trees. For the rest of the year decision trees accuracies are superior. It can be observed again that in hot months of the year neural network performs better than decision trees.

Table 7: MMAPE of Metro East out of sample data (2010 as the test year).

	Daily MMAPE NN	Daily MMAPE DT	Weekly MMAPE NN	Weekly MMAPE DT	Average Temperature (°C)
Jan	2.8%	3.4%	3.9%	4.1%	25.1
Feb	2.3%	2.4%	3.5%	3.7%	24.4
Mar	2.6%	2.4%	3.6%	3.5%	23.2
Apr	2.2%	1.4%	3.3%	2.8%	19.1
May	2.3%	1.4%	3.3%	2.9%	14.6
Jun	2.1%	1.5%	2.9%	2.7%	12.0
Jul	2.3%	1.8%	3.1%	2.9%	11.3
Aug	2.3%	1.6%	3.3%	2.7%	12.3
Sep	2.1%	1.4%	3.0%	2.8%	14.8
Oct	1.9%	1.3%	3.0%	2.5%	17.3
Nov	2.1%	2.5%	3.3%	3.8%	22.0
Dec	2.3%	2.3%	3.3%	3.6%	22.5

Table 8 belongs to Metro South where consumers are a combination of residential and industrial ones. All of the daily MMAPE values are below 4%. Similar to previous ones for hotter months of the year (December to March) neural network performs better comparing to decision trees. For the rest of the year decision trees accuracies are superior.

Table 8: MMAPE of Metro South out of sample data (2010 as the test year).

	Daily MMAPE NN	Daily MMAPE DT	Weekly MMAPE NN	Weekly MMAPE DT	Average Temperature (°C)
Jan	3.8%	4.0%	5.0%	5.3%	25.5
Feb	3.5%	4.0%	4.4%	5.3%	24.4
Mar	3.2%	3.9%	4.5%	4.9%	23.3
Apr	2.4%	2.0%	3.2%	3.1%	18.7
May	1.8%	1.6%	3.2%	2.9%	14.7
Jun	2.7%	2.1%	3.5%	3.2%	12.3
Jul	3.3%	3.0%	4.1%	4.0%	11.3
Aug	3.0%	2.5%	4.2%	4.0%	12.3
Sep	2.7%	1.9%	3.8%	3.5%	14.8
Oct	2.4%	1.4%	3.6%	2.9%	17.4
Nov	3.3%	3.1%	4.5%	4.0%	22.0
Dec	3.5%	4.0%	4.7%	5.3%	22.5

Table 9 represents the MMAPE values for the CBD region which is mainly composed of commercial consumers. All of the daily MMAPE values are below 5%. Similar to previous ones for hotter months of the year (January to March) neural network performs better compared to decision trees. For the rest of the year decision trees accuracies are superior.

Table 9: MMAPE of CBD out of sample data (2010 as the test year).

	Daily MMAPE NN	Daily MMAPE DT	Weekly MMAPE NN	Weekly MMAPE DT	Average Temperature (°C)
Jan	4.3%	4.9%	5.5%	5.8%	25.5
Feb	3.7%	4.0%	4.7%	4.9%	24.4
Mar	4.1%	4.5%	5.4%	5.6%	23.1
Apr	3.0%	3.0%	4.0%	3.8%	18.7
May	2.0%	1.4%	2.8%	2.3%	14.7
Jun	2.5%	2.4%	3.4%	3.3%	12.3
Jul	2.6%	2.0%	3.7%	3.0%	11.3
Aug	2.6%	1.4%	3.8%	2.3%	12.3
Sep	2.9%	1.6%	3.8%	2.5%	14.8
Oct	2.7%	1.5%	3.8%	2.5%	17.4
Nov	3.3%	2.8%	4.3%	4.0%	22.0
Dec	4.7%	4.2%	5.6%	5.1%	22.5

Table 10 contains the MMAPE values of neural networks and decision trees for Country North region. All daily MMAPE values are below 5%. Interestingly for this region decision trees perform better than neural network for the whole length of the year and irrespective of the ambient temperature. Majority of consumers of Country North are industrial.

Table 10: MMAPE of Country North out of sample data (2010 as the test year).

	Daily MMAPE NN	Daily MMAPE DT	Weekly MMAPE NN	Weekly MMAPE DT	Average Temperature (°C)
Jan	4.2%	3.5%	5.3%	4.2%	24.6
Feb	3.8%	3.3%	5.3%	5.0%	25.1
Mar	4.7%	4.3%	5.8%	5.3%	24.4
Apr	2.7%	2.3%	3.8%	3.5%	20.3
May	3.9%	3.3%	4.8%	4.3%	16.9
Jun	3.7%	3.1%	4.5%	4.2%	14.5
Jul	3.7%	3.3%	4.8%	4.5%	12.2
Aug	3.3%	3.1%	4.5%	4.4%	13.4
Sep	3.1%	3.0%	4.9%	4.5%	15.3
Oct	4.2%	3.9%	6.2%	5.3%	18.9
Nov	3.9%	3.3%	5.2%	4.8%	22.9
Dec	3.5%	3.0%	4.5%	4.2%	23.2

Table 11 contains the MMAPE values of neural networks and decision trees for Country East region. All daily MMAPE values are below 5%. For this case decision trees perform better than neural network for the whole length of the year and irrespective of the ambient temperature. Majority of consumers of this region are industrial.

Table 11: MMAPE of Country East out of sample data (2010 as the test year).

	Daily MMAPE NN	Daily MMAPE DT	Weekly MMAPE NN	Weekly MMAPE DT	Average Temperature (°C)
Jan	3.6%	3.3%	4.4%	4.1%	26.7
Feb	4.2%	3.8%	5.5%	4.9%	25.2
Mar	4.0%	3.8%	5.1%	5.0%	25.6
Apr	4.1%	3.5%	5.2%	4.8%	25.7
May	4.0%	3.5%	5.2%	4.6%	24.6
Jun	4.3%	3.0%	5.2%	4.2%	12.2
Jul	4.3%	3.7%	5.5%	4.5%	10.9
Aug	3.8%	2.6%	4.7%	3.8%	12.1
Sep	3.7%	2.6%	4.9%	3.5%	14.4
Oct	3.5	3.3	4.8	4.5	18.0
Nov	3.4%	3.1%	4.4%	4.2%	22.6
Dec	3.1%	3.0%	4.2%	3.8%	23.3

Table 12contains the MMAPE values of neural networks and decision trees for Country South region. All daily MMAPE values are below 5%. For this case decision trees perform better than neural network for the whole length of the year and irrespective of the ambient temperature. Majority of consumers of this region are industrial.

Table 12: MMAPE of Country South out of sample data (2010 as the test year).

	Daily MMAPE NN	Daily MMAPE DT	Weekly MMAPE NN	Weekly MMAPE DT	Average Temperature (°C)
Jan	4.3%	2.4%	5.8%	3.9%	21.7
Feb	3.6%	2.7%	4.8%	3.6%	21.9
Mar	2.9%	2.7%	4.4%	4.1%	20.9
Apr	2.2%	2.1%	3.4%	3.0%	17.5
May	3.1%	2.4%	4.3%	3.5%	14.7
Jun	3.0%	2.7%	4.0%	3.9%	12.7
Jul	2.8%	2.5%	4.0%	3.6%	11.9
Aug	3.3%	2.9%	4.5%	3.7%	12.2
Sep	3.3%	2.9%	4.4%	4.2%	13.7
Oct	2.8%	2.3%	3.5%	3.1%	15.5
Nov	4.3%	3.8%	5.7%	4.8%	19.9
Dec	3.4%	3.4%	4.8%	4.5%	19.6

Table 13contains the MMAPE values of neural networks and decision trees for Country Goldfields region. All daily MMAPE values are below 5%. For this case decision trees perform better than neural network for the whole length of the year and irrespective of the ambient temperature. Majority of consumers of this region are industrial.

Table 13: MMAPE of Country Goldfields out of sample data (2010 as the test year).

	Daily MMAPE NN	Daily MMAPE DT	Weekly MMAPE NN	Weekly MMAPE DT	Average Temperature (°C)
Jan	3.3%	3.0%	4.1%	3.7%	26.2
Feb	3.1%	2.6%	4.1%	3.5%	24.4
Mar	3.0%	2.3%	4.0%	3.2%	20.9
Apr	3.1%	2.6%	4.0%	3.4%	16.8
May	3.9%	2.1%	4.8%	3.1%	13.0
Jun	4.5%	2.3%	6.5%	3.2%	11.1
Jul	3.6%	2.3%	4.7%	3.9%	11.1
Aug	3.9%	2.4%	5.6%	3.3%	13.1
Sep	3.3%	2.6%	4.2%	3.8%	16.2
Oct	3.8%	2.0%	4.8%	3.0%	20.4
Nov	4.1%	2.5%	5.3%	3.1%	23.7
Dec	3.7%	2.5%	4.8%	3.6%	23

5.7 CONCLUSION

[123] performs extensive studies on accuracies of electricity forecasting methods, and discusses the significance of accurate forecasts and their effects on electricity market. Interestingly, it concludes that the MMAPE of less than 5% is within the range of adequate forecast. In simple words, he demonstrates that accuracies of less than 5% have minimal impact on economics of the electricity market. Based on this research the presented MMAPEs in Table 6 to Table 13 are all considered as accurate.

It is also observed that during the hot months neural networks are a better choice for forecasting the residential and commercial electricity load, and for the remainder of the year decision trees are better. This differs for the industrial load. It can be seen that decision trees are a better choice for all the months of the year irrespective of the ambient temperature. Decision trees are better than neural networks when the system nonlinearity is low. However,

when the system nonlinearity increases neural networks are better. In load forecasting applications, the higher the temperature sensitivity, the higher nonlinearities will be in the system. This is because the behaviour of consumers change with weather. They decide whether to turn on-off sources of cooling or heating.

It can be concluded that both methods are capable of forecasting the electricity load with high accuracy, but one may perform better than the other, depending on the characteristics of the case study. Using the load type determination criterion will help the planner to extract the dominant component of the electricity load and decide which method to use. This research suggests that bagging decision trees for industrial loads, and, based on the temperature sensitivity of the system, decision trees or a combination of decision trees and neural networks can be used. A similar approach can be applied to other forecasting methods to identify the best possible method in different conditions. Overall, a system based approach can be much more reliable than a classic approach.

Here is an example to further clarify the benefits of a combined system based approach.

By looking at the CBD region QQ plots (Figure 63 to Figure 65), a data analyst can conclude that the electricity consumers of that region are mainly commercial ones. Based on this conclusion he/she will be able touse neural networks to forecast the electricity load on the hot months of year (which are January, February and March for this case), and decision trees to be used for the rest of the year. See Figure 66 to see what he/she gets as the result, compared to using only one model.

Blue bars represent the daily MMAPE of CBD region for 2010 as test year, using the system based combined approach. Red bars and green bars respectively represent the daily MMAPE percentage of the same region using neural networks and decision trees. System based combined approach guarantees the best result all year round.

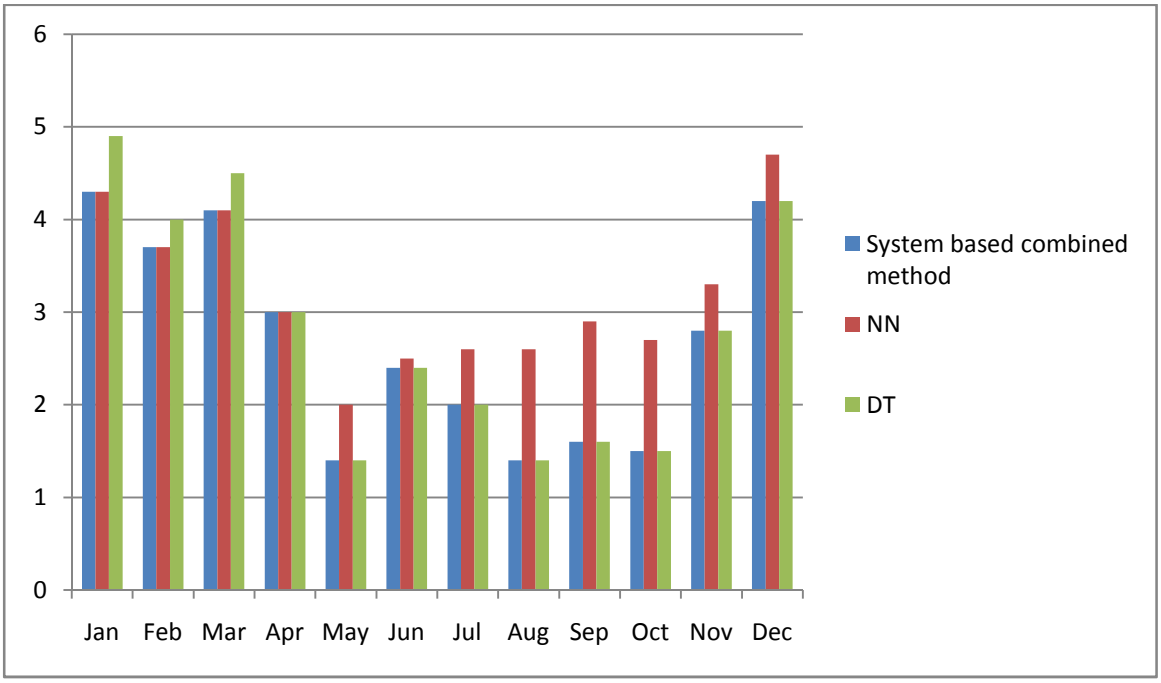


Figure 66: CBD MMAPE% versus month of the year for system based combined method, Neural networks and decision trees (forecast for 2010 as the test year)

Chapter

6 MEDIUM-TERM LOAD FORECASTER (SYSTEM BASED APPROACH)

6.1 ABSTRACT

A system based approach for short-term load forecasting (STLF) of the SWIS was presented in the previous chapter. This chapter will present a similar approach to solve the problem of medium-term load forecasting (MTLF). Two main differences between STLF and MTLF are the availability of weather data and the forecasting objectives. Two methods are applied, namely, neural networks and regression trees, and the results are proven to be satisfactory for one year ahead spatial medium forecasts.

6.2 INTRODUCTION

One of the main differences between STLF and MTLF is the availability of the information. The temperature forecasts of up to seven days, which play a significant role in STLF applications, are not available for MTLF. The reason is that the complexity and random behaviour of climatic systems make it unrealistic to forecast the weather information for a couple of months ahead. Another major difference between STLF and MTLF is their application in the electricity industry. Rather than looking for the amount of electricity consumed at each point in time, like STLF, planners are keen to have the information of peak load and energy consumption in medium-term forecasts[124].

6.3 INPUT DATA PREPARATION

6.3.1 EXTRACTING PEAK LOAD AND ENERGY CONSUMPTION DATA

As discussed in previous chapters, the available electricity load data is composed of half an hourly consumption data of eight regions. For medium-term load forecasting applications, peak load and energy consumption are required instead of half an hourly load data[125]. A function is coded to extract energy consumption in kilo watt hour and peak load in kilo watt from original load data matrices.[34] This involves rearranging the available data in monthly format and chronological order. This has been applied on all eight regions of the SWIS and the results are presented in Figure 67 to Figure 74.The left panels present the extracted data in 168 consecutive months[35], and the right panels illustrate the energy consumption and peak load data behaviour of each month over 14 years.

Figure 67 illustrates the prepared load data for CBD region. Seasonality in peak load consumption is very obvious by looking at the top left pane. There is also a constant rise of peak values over time. The bottom left pane represents the energy consumption for the same region.

[34] See Appendix D for the codes.
[35] From January 1996 to December 2009

Seasonality in the data is less obvious. There is a constant rise in consumption data during the course of fourteen years. The right panes show peak load and energy consumption of CBD region with a coloured line dedicated to each month of the year. Because some seasonality information is related to month of the year much less seasonality is present in these data sets. A constant growth is visible in all the values. This shows that CBD region have been under constant expansion[36]during those fourteen years.

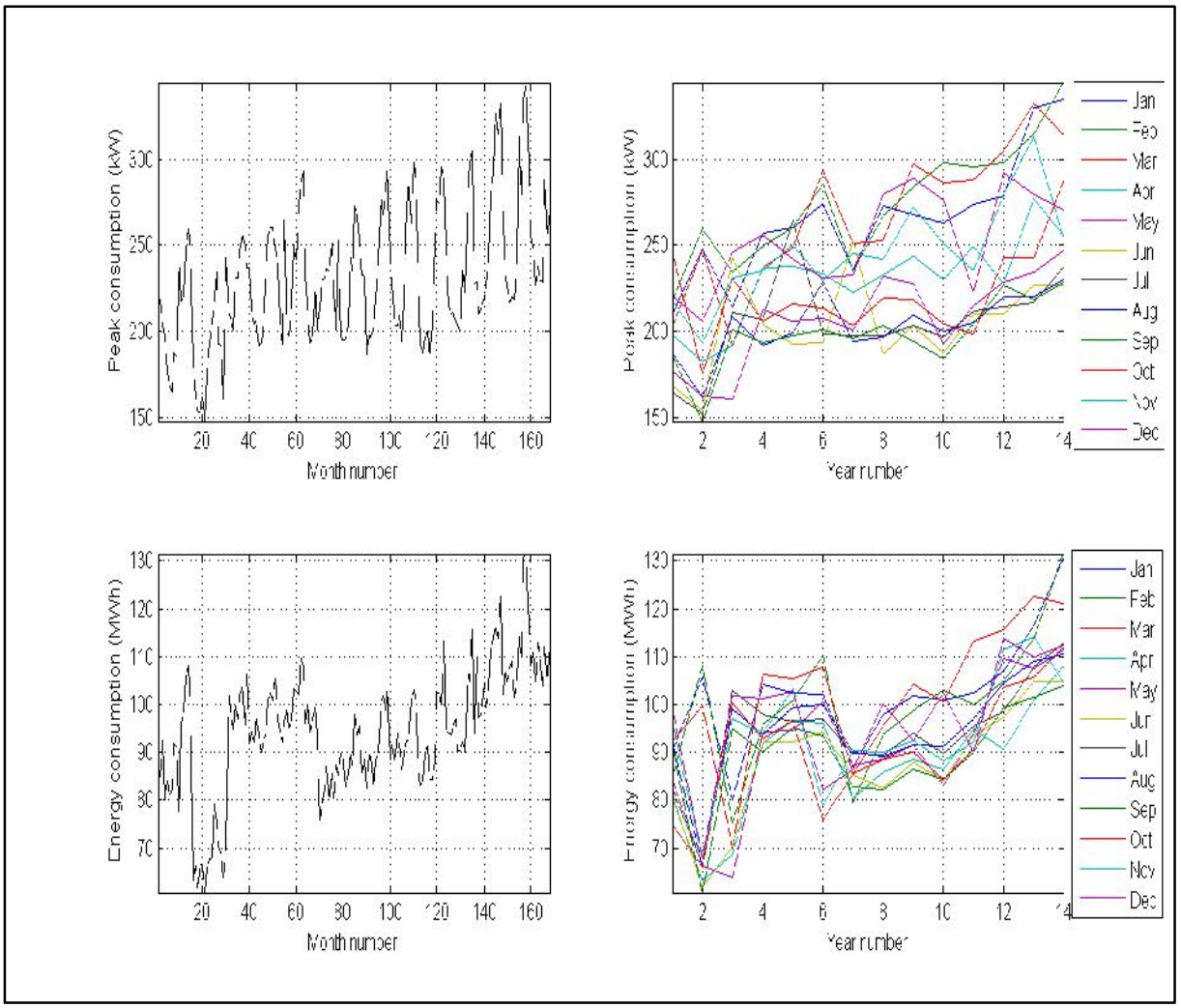


Figure 67: CBD load data preparation.

Figure 68 illustrates the prepared load data for Country East region. Seasonality is very obvious by looking at the top left pane. There is a sharp rise in the first three years, and then there is a drop and a constant trend for the remaining years. Any rise in electricity demand figures in an evidence of expansion. A drop in a graph of an industrially dominated region like this is an evidence of plant decommissioning or economy down turn. The bottom left pane represents the energy consumption for the same region. Seasonality in the data is less obvious. There is sharp

[36] The term expansion here is referring to the growth in population, buildings and infrastructure which results in higher electricity demand.

rise in the first three years and a constant trend for the remaining years. Same sort of increasing trend for the first three years is present in the right panes.

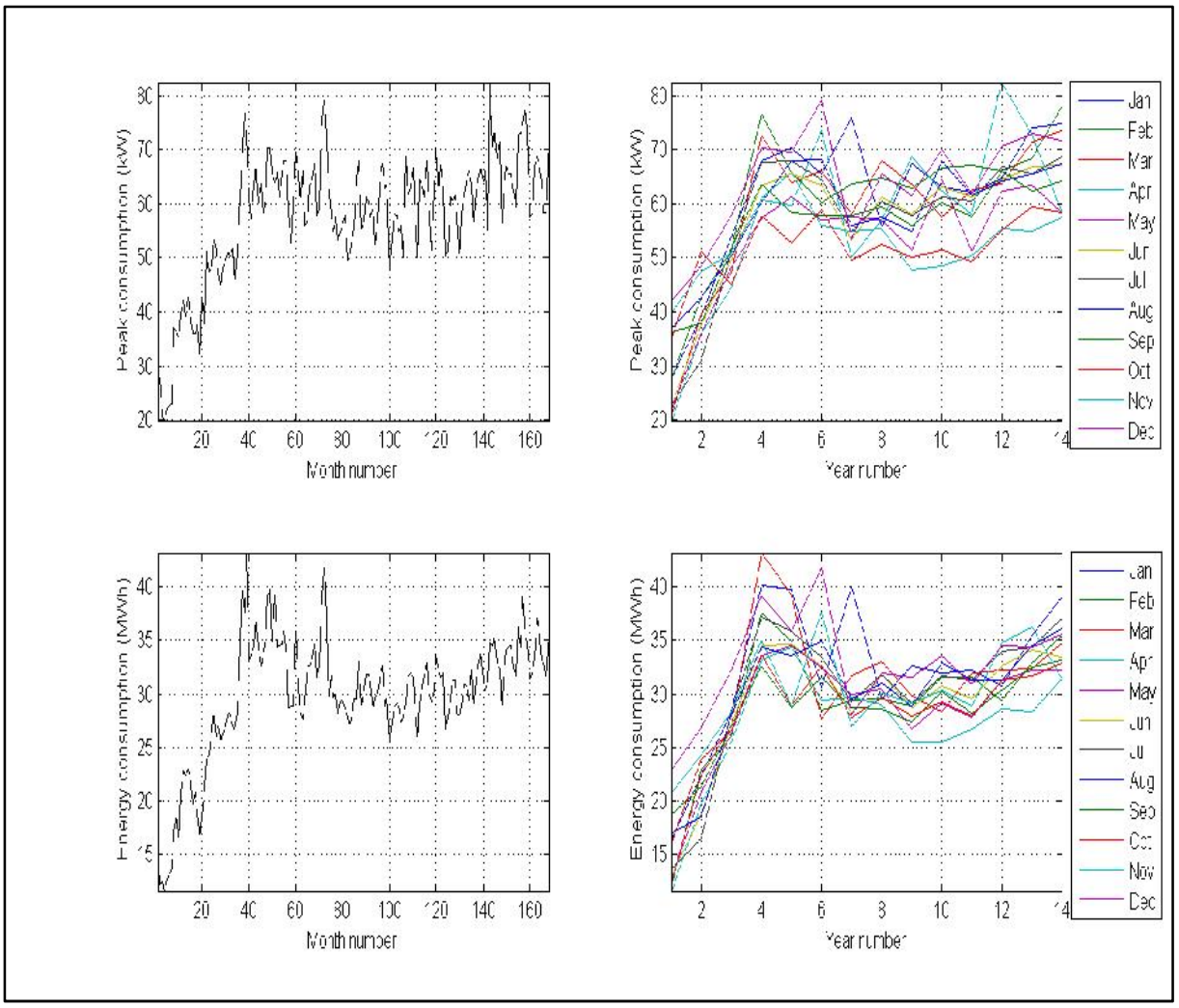


Figure 68: Country East load data preparation.

Figure 69illustrates the prepared load data for Country Goldfields region. There is sharp rise in the first three years and almost a constant trend for the remaining years. At year ten there is a drop in the peak load electricity in the region. The bottom left pane represents the energy consumption for the same region. Seasonality in the data is less obvious. There is sharp rise in the first three years and a constant trend for the remaining years. At year ten there is a drop in the electricity consumption in the region. Same sort of trend exist on the right panes. As stated earlier, the drop happening at around year ten could correlate with a decommissioning of a major plant in the region or economy downturn.

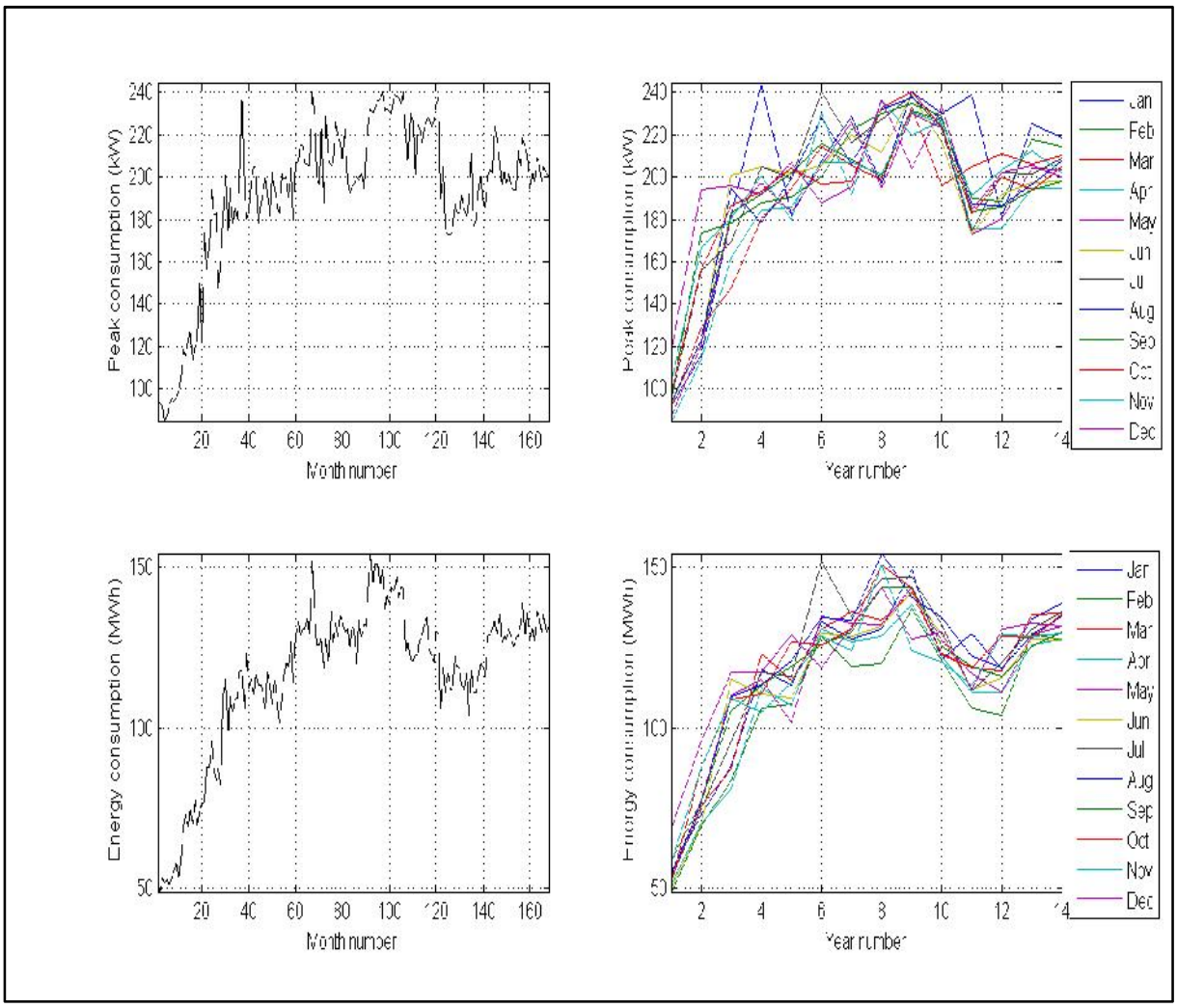


Figure 69: Country Goldfields load data preparation.

Figure 70 illustrates the prepared load data for Country North region. There is a small drop in the peak load in the first three years which is then followed up by a constant increasing trends for the following years. The bottom left pane represents the energy consumption for the same region. Seasonality in the data is less obvious. There is a small drop in the peak load in the first three years which is then followed up by a constant increasing trends for the following years. The right panes show peak load and energy consumption of Country North region with a coloured line dedicated to each month of the year. The same sort of trend is visible in monthly data. It could be because of a plant decommissioning or overhaul on the first three years and constant industrial expansion of the area for the rest of the years.

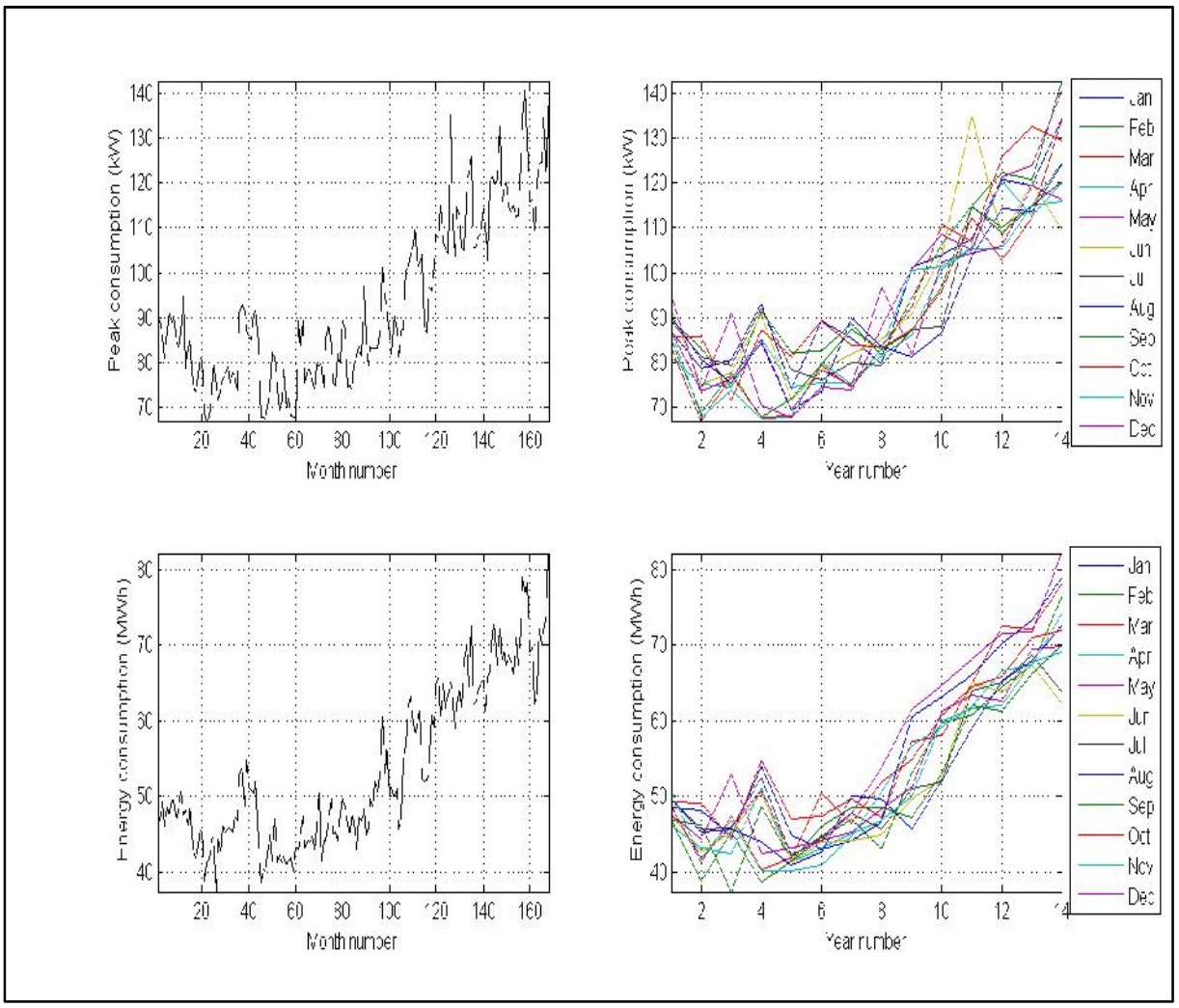


Figure 70: Country North load data preparation.

Figure 71illustrates the prepared load data for Country South region. There is sharp rise in the first three years and almost a constant trend for the remaining years. This could relate to industrial expansions happened in the first three years.

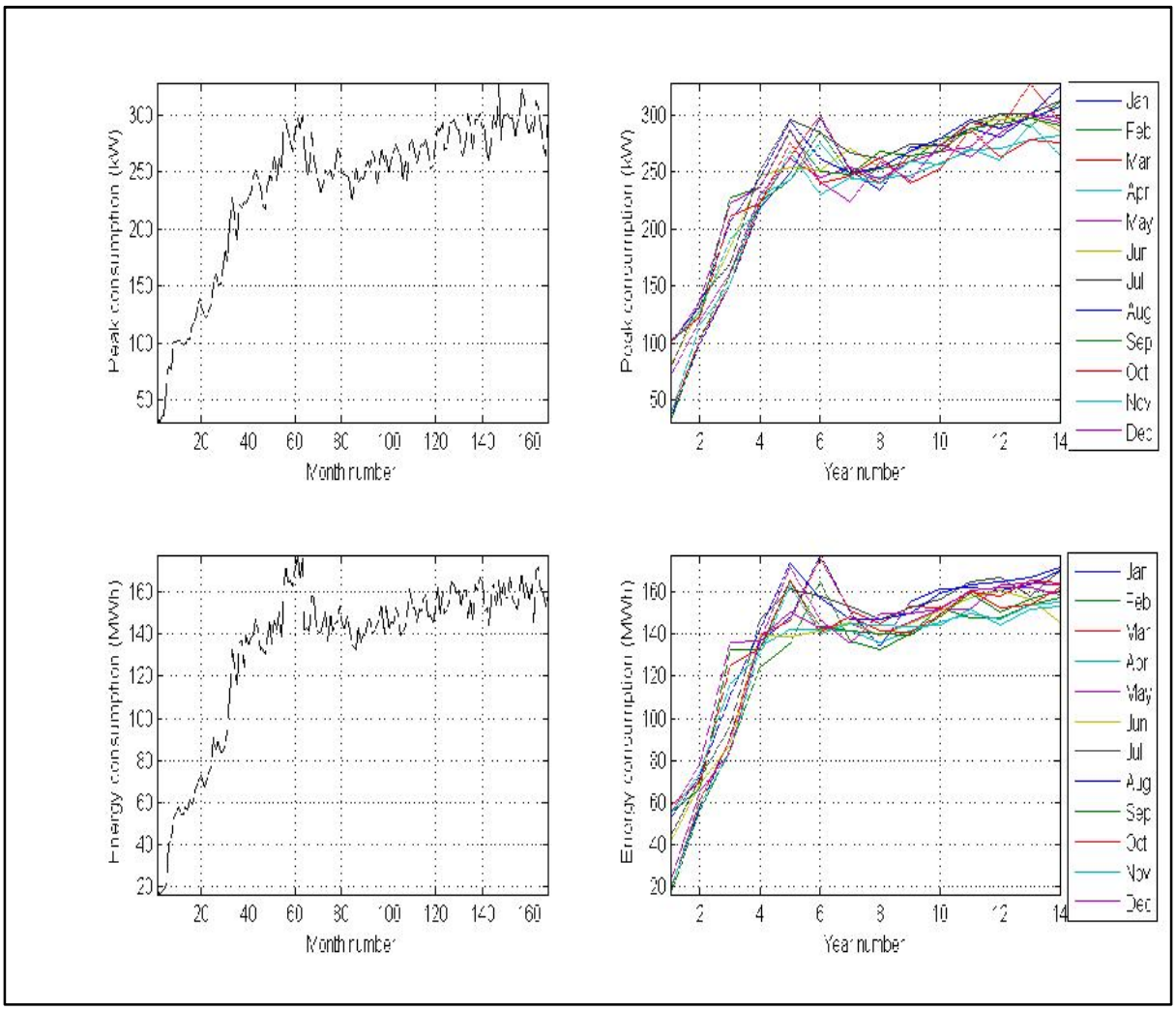


Figure 71: Country South load data preparation.

Figure 72illustrates the prepared load data for Metro East region. Seasonality is very obvious by looking at the graph. There is also a constant rise of peak and energy consumption values over time. This is in line with constant expansion of the Metro East region during those fourteen years.

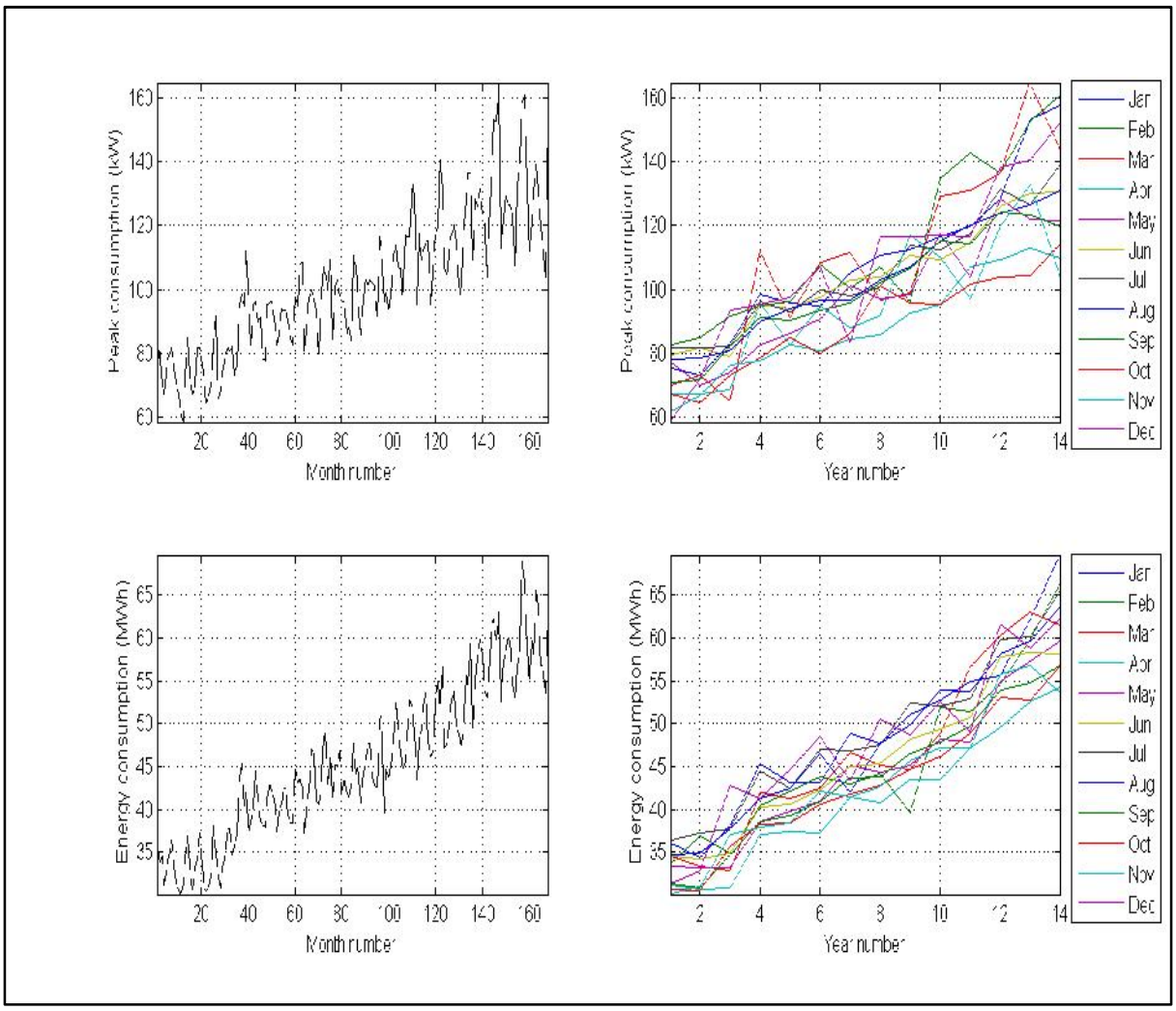


Figure 72: Metro East load data preparation.

Figure 73illustrates the prepared load data for Metro North region. The top left panel shows the peak load data for 168 consecutive months. Seasonality is very obvious by looking at the graph. There is also a constant rise of peak values over time. The bottom left pane represents the energy consumption for the same region. Seasonality is very obvious by looking at the graph. There is a constant rise in consumption data during the course of fourteen years. The right panes show peak load and energy consumption of Metro North region with a coloured line dedicated to each month of the year. Because some seasonality information is related to month of the year much less seasonality is present in these data sets. A constant growth is visible in all the values. This is in line with constant expansion of the Metro North region during those fourteen years.

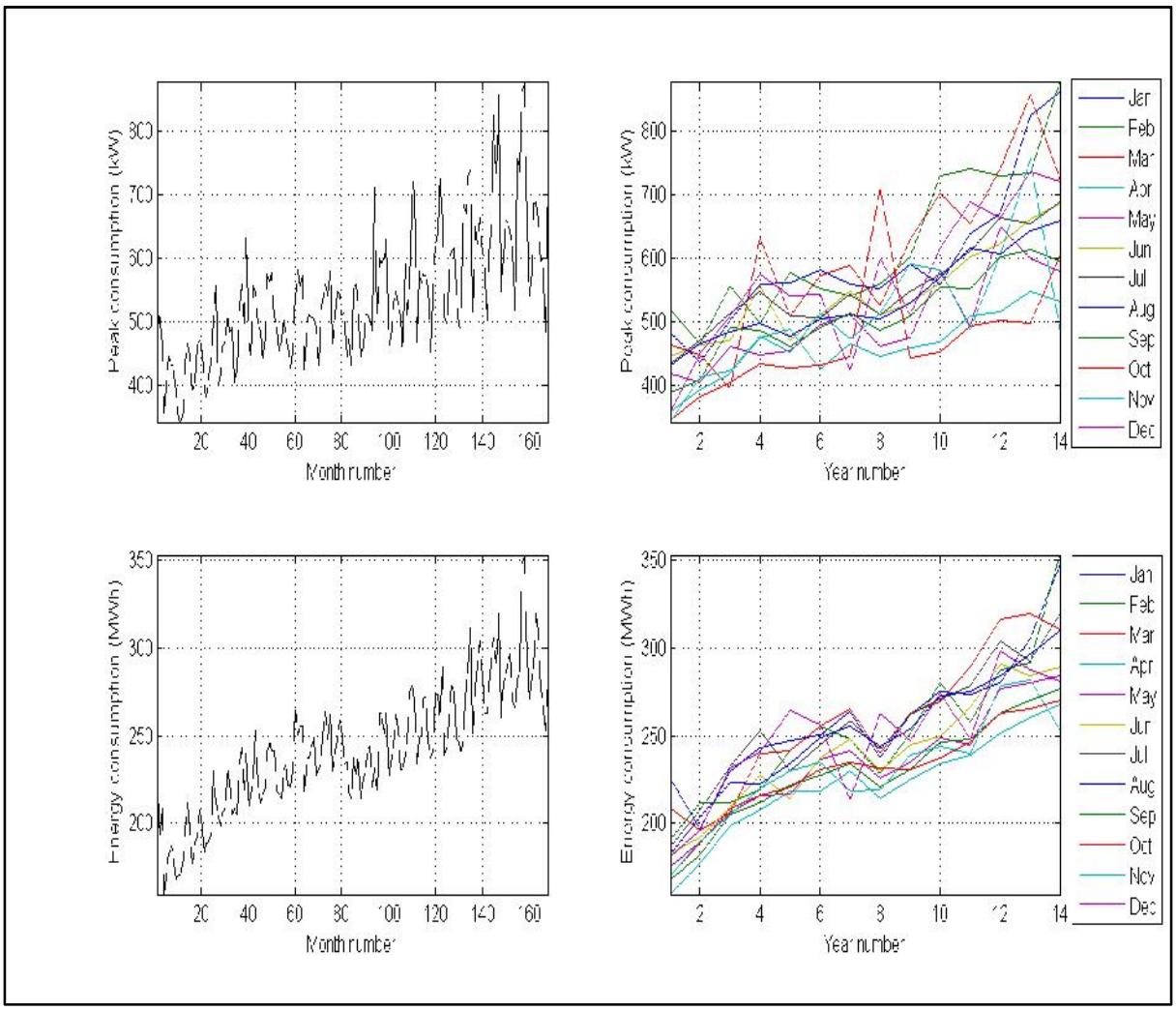


Figure 73: Metro North load data preparation.

Figure 74illustrates the prepared load data for Metro South region. The top left panel shows the peak load data for 168 consecutive months. Seasonality is very obvious by looking at the graph. There is also a constant rise of peak values over time. The bottom left pane represents the energy consumption for the same region. Seasonality is very obvious by looking at the graph. There is a constant rise in consumption data during the course of fourteen years. The right panes show peak load and energy consumption of Metro South region with a coloured line dedicated to each month of the year. Because some seasonality information is related to month of the year much less seasonality is present in these data sets. A constant growth is visible in all the values. This is in line with constant expansion of the Metro South region during those fourteen years.

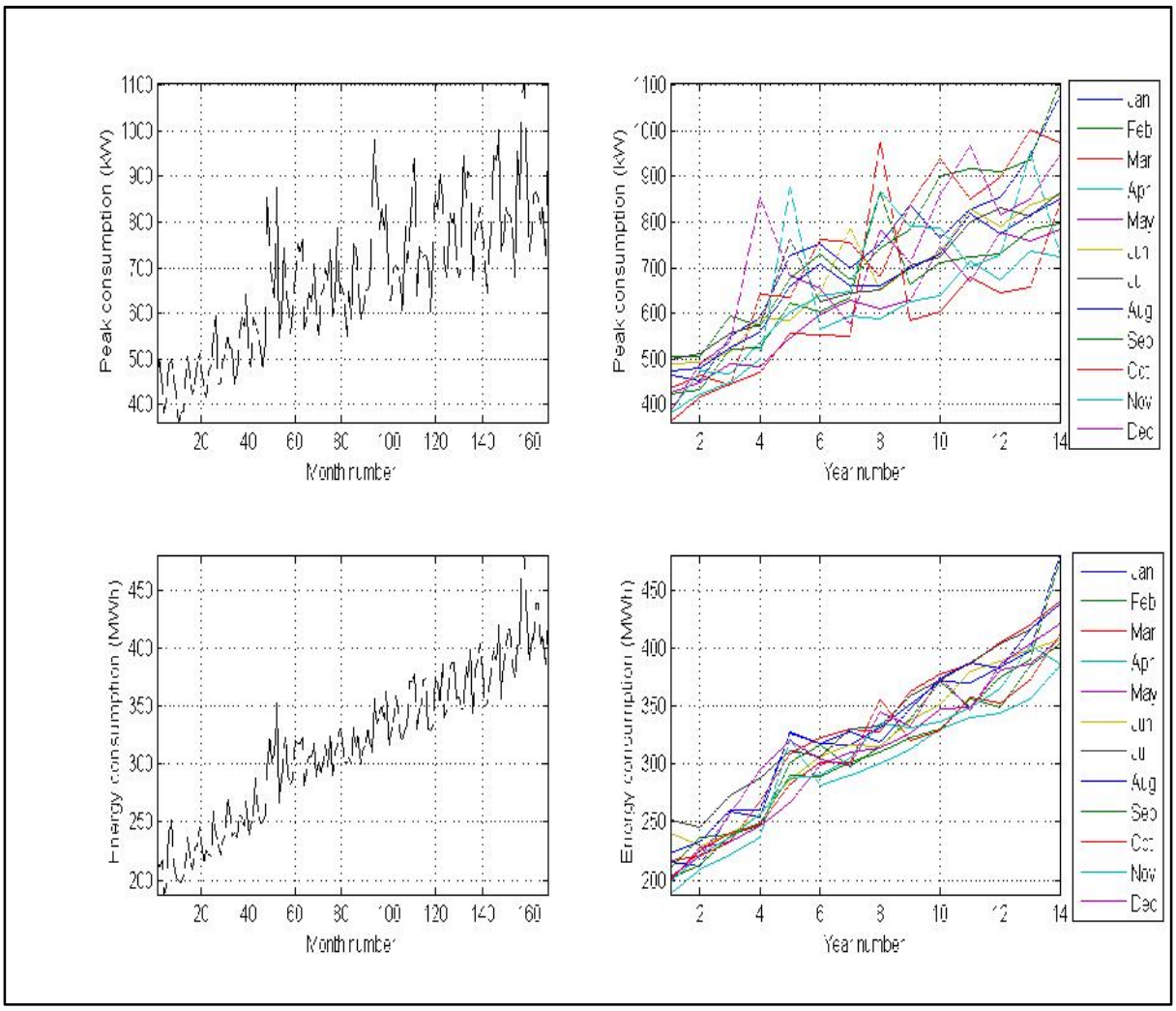


Figure 74: Metro South load data preparation.

Based on the above observations regularities such as seasonality and trends are obvious on all the peak load and energy consumption graphs.

The seasonality is when the data set repeats itself in constant time intervals. It can be daily, weekly, monthly or yearly. Daily seasonality directly relates to regular daily activities at offices, houses and industrial facilities. Weekly seasonality happens mainly because different sort of activities are being performed during weekends compared to the week days. For industrial consumers who work the same schedule around the week, this type of seasonality won't be present. Monthly seasonality is to do with temperature change over the period of each month. As studied earlier temperature changes play an important role on the amount of electricity being consumed. And finally, yearly seasonality is mainly to do with public holidays and temperature.

Trends in the data are the evidence of constant load change over a period of time. This could be because of population increase in residential and commercial areas. Trends are also visible in

industrial loads where the industrial facilities go through expansion, decommissioning or overhaul.

6.3.1 DEALING WITH TRENDS AND SEASONALITY

Irrespective of the training model that it is being used for load forecasting, it is always a good practice to remove trends and seasonality from the data set before feeding them to the training model. If data with obvious seasonality and trend is being used as the input, the training model will capture those obvious trends and seasonality instead of focusing on capturing more random behaviours[37] of the load data consumers. To avoid this, it was decided to deal with trends and seasonality separately. In simple words the trends and seasonality are being modelled separately and get deducted from the load data. The remaining values or residuals will be used in training models to capture random behaviours.

The first step involved eliminating the last 12 samples of datasets, as the test year and the remainder of the data are analysed in a search for the best detrending methodology. Given the visual characteristics of the graphs, polynomial trend was used to detrend the data. Several polynomials were fitted to the training data, and the goodness of fit, residuals bar plot, and the norm of the residuals are presented in Figure 75to Figure 90.

Figure 75 represents the energy consumption of the CBD region for thirteen years. The year number fourteen is taken out of the data. It will be used as the test year to evaluate the model. The top pane shows the data set with 4th, 5th and 6th degree polynomial fits. The bottom pane shows the norm of residuals of each fit. The lower the norm of the residuals, the better the fit will be. In this case the residuals are 124.5502, 116.2753, 114.4536 respectively for 4th, 5th and 6th degree polynomial fits. The 6th degree polynomial fit with the lowest residual is showing the best fit for the CBD energy consumption.

[37]Random behaviours are the same as residuals

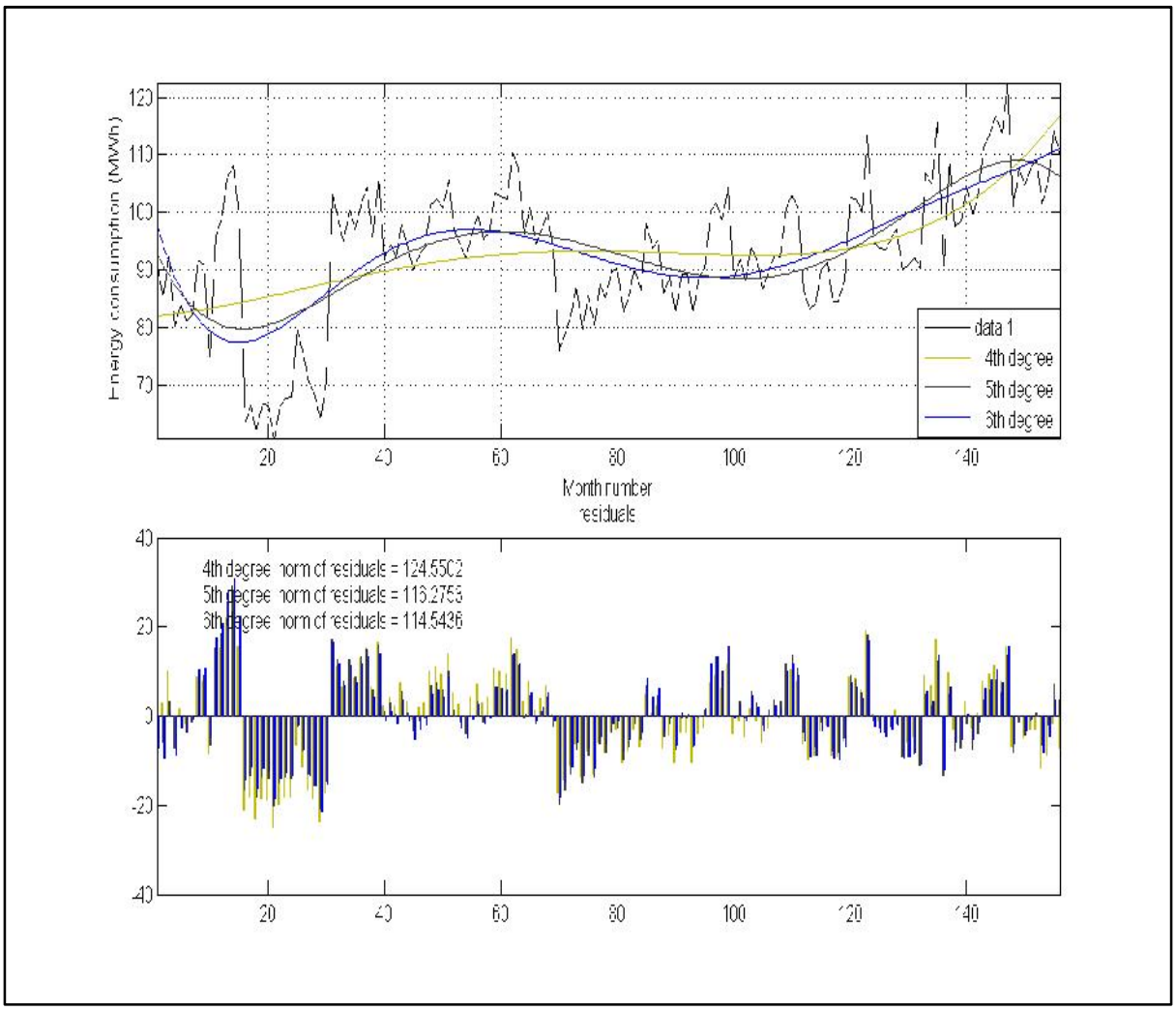


Figure 75: CBD energy consumption detrending.

Figure 76 represents the peak consumption of the CBD region for thirteen years. The year number fourteen is taken out of the data. It will be used as the test year to evaluate the model. The top pane shows the data set with 4th, 5th and 6th degree polynomial fits. The bottom pane shows the norm of residuals of each fit. All curves seems to be behaving very similar in-terms of capturing the data trend. The lower the norm of the residuals the better the fit will be. In this case the residuals are 415.3223, 413.4267, 412.8251, respectively for 4th, 5th and 6th degree polynomial fits. The 6th degree polynomial fit with the lowest residual is showing the best fit for the CBD peak consumption.

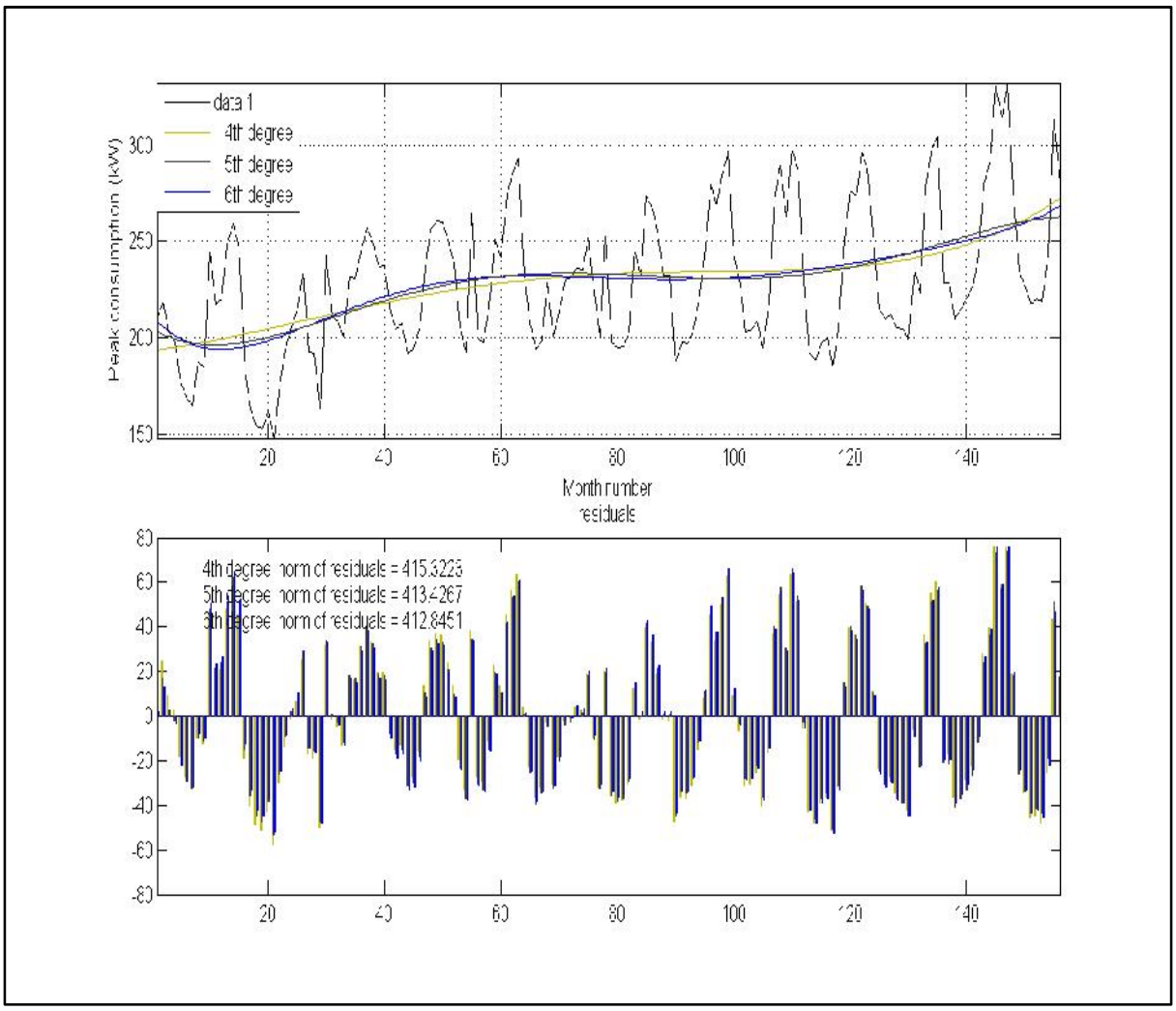


Figure 76: CBD peak consumption detrending.

Figure 77 represents the energy consumption of the Country North region for thirteen years. The year number fourteen is taken out of the data. It will be used as the test year to evaluate the model. The top pane shows the data set with 4th, 5th and 6th degree polynomial fits. The bottom pane shows the norm of residuals of each fit. All curves seems to be behaving very similar in-terms of capturing the data trend. The lower the norm of the residuals, the better the fit will be. In this case the residuals are 44.597, 44.4497, 43.8084, respectively for 4th, 5th and 6th degree polynomial fits. The 6th degree polynomial fit with the lowest residual is showing the best fit for the Country North energy consumption.

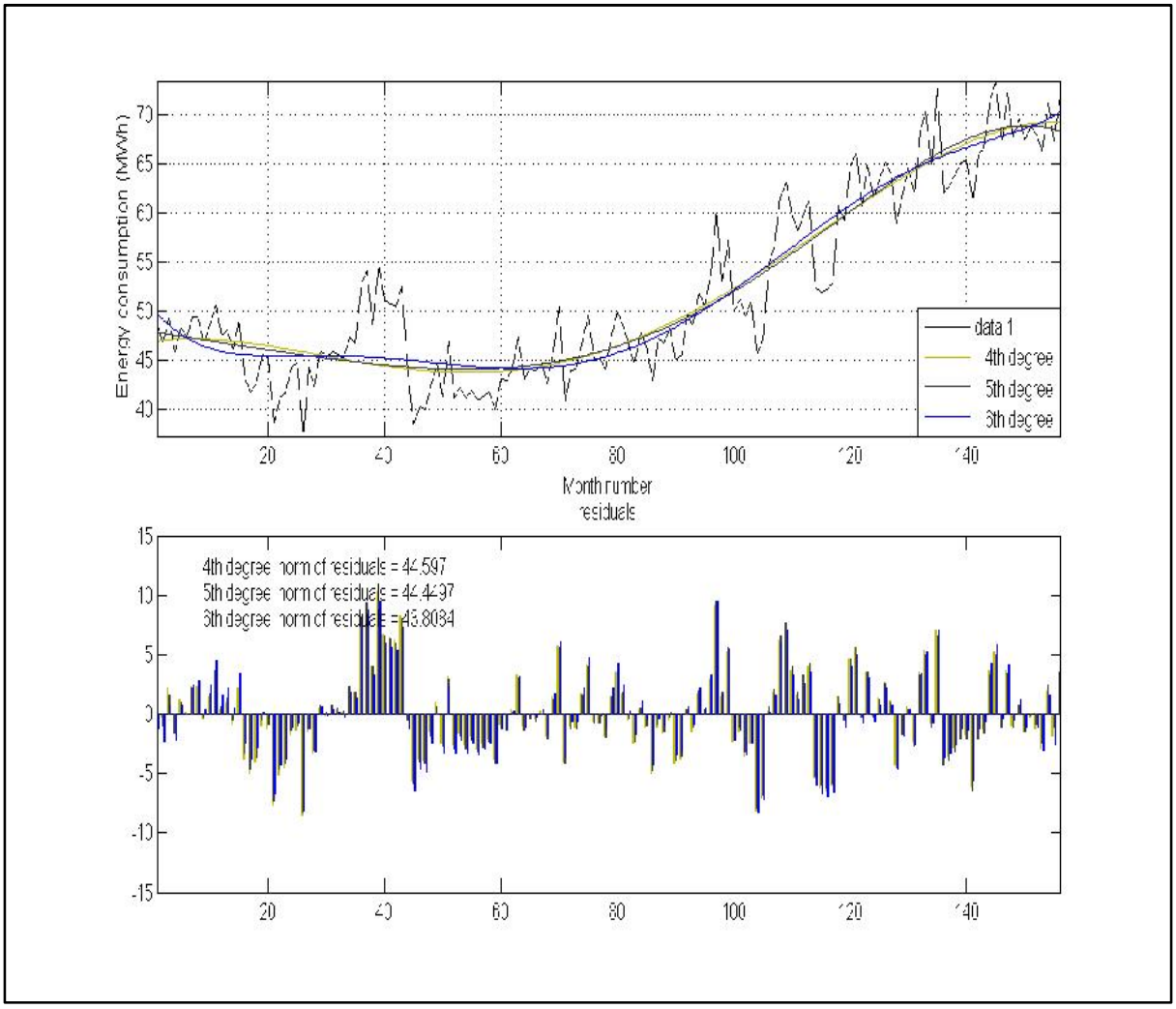


Figure 77: Country North energy consumption detrending.

Figure 78 represents the peak consumption of the Country North region for thirteen years. The year number fourteen is taken out of the data. It will be used as the test year to evaluate the model. The top pane shows the data set with 4th, 5th and 6th degree polynomial fits. The bottom pane shows the norm of residuals of each fit. All curves seems to be behaving very similar in-terms of capturing the data trend. The lower the norm of the residuals the better the fit will be. In this case the residuals are 89.8812, 88.6282, 88.0803, respectively for 4th, 5th and 6th degree polynomial fits. The 6th degree polynomial fit with the lowest residual is showing the best fit for the Country North peak consumption.

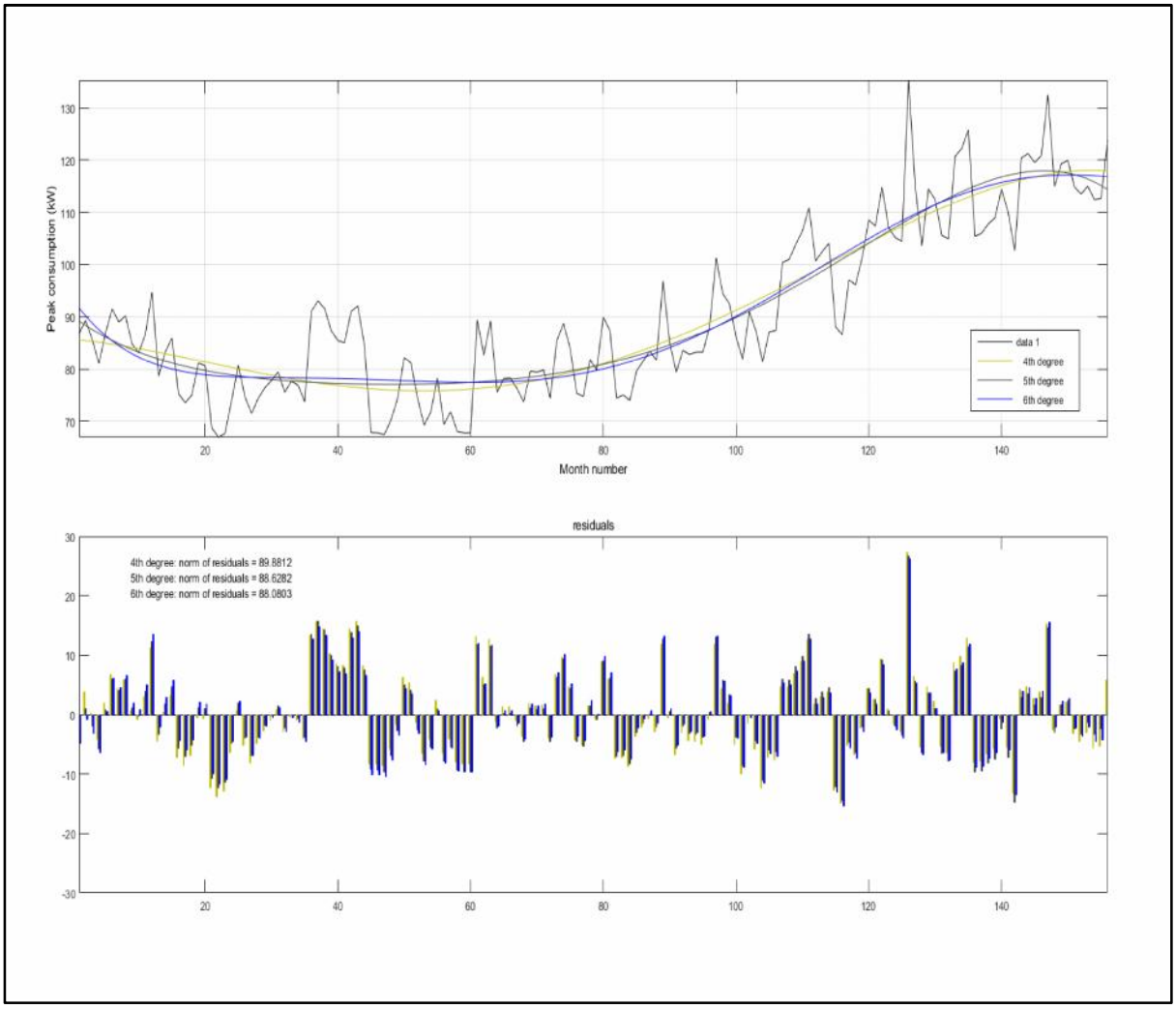


Figure 78: Country North peak detrending.

Figure 79 represents the energy consumption of the Country East region for thirteen years. The year number fourteen is taken out of the data. It will be used as the test year to evaluate the model. The top pane shows the data set with 4th, 5th and 6th degree polynomial fits. The bottom pane shows the norm of residuals of each fit. All curves seem to be behaving very similar in-terms of capturing the data trend. The lower the norm of the residuals the better the fit will be. In this case the residuals are 37.6567, 36.4615, 35.1391, respectively for 4th, 5th and 6th degree polynomial fits. The 6th degree polynomial fit with the lowest residual is showing the best fit for the Country East energy consumption.

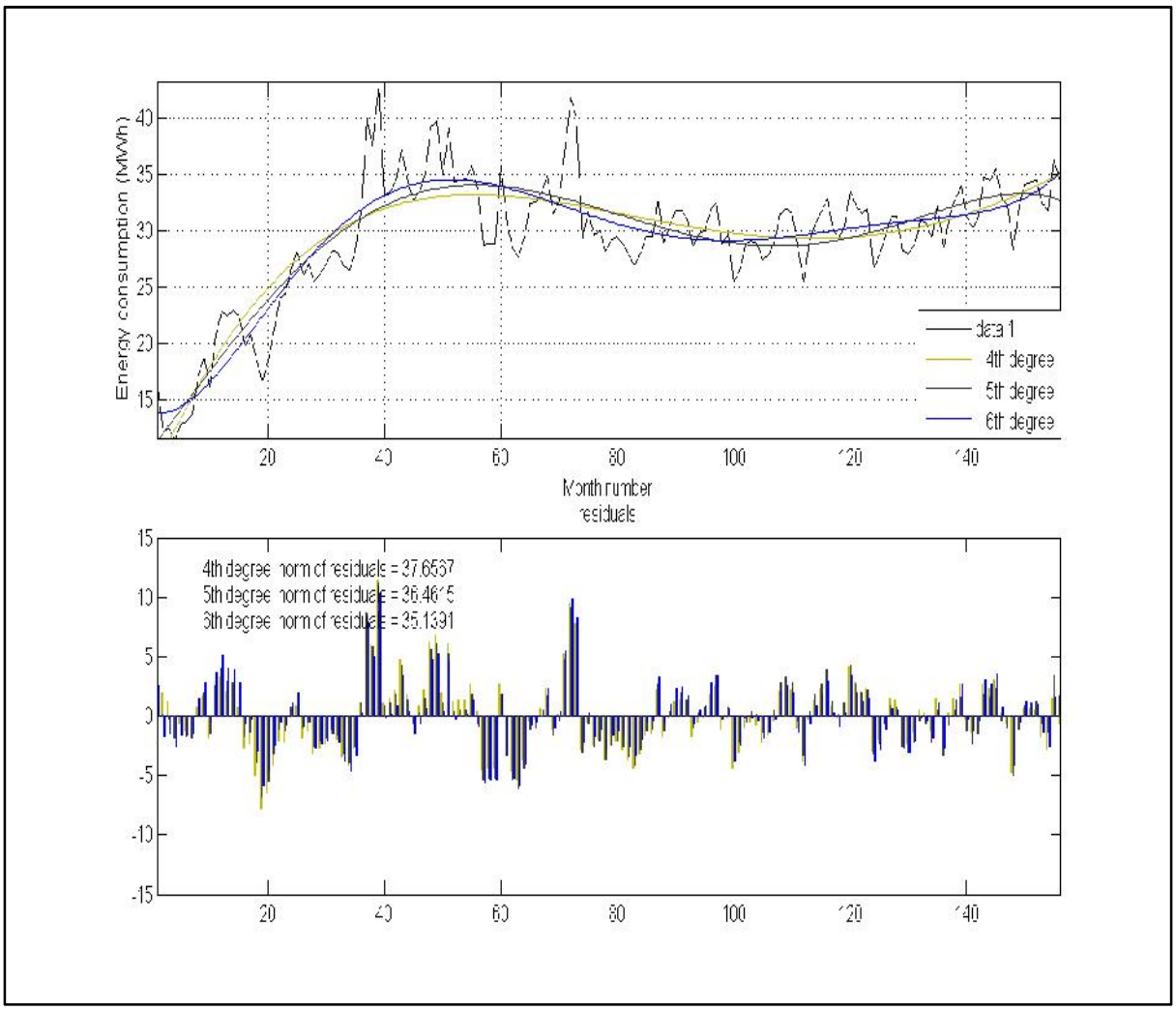


Figure 79: Country East energy consumption detrending.

Figure 80 represents the peak consumption of the Country East region for thirteen years. The year number fourteen is taken out of the data. It will be used as the test year to evaluate the model. The top pane shows the data set with 4th, 5th and 6th degree polynomial fits. The bottom pane shows the norm of residuals of each fit. All curves seem to be behaving very similar in-terms of capturing the data trend. The lower the norm of the residuals the better the fit will be. In this case the residuals are 37.6567, 36.4615, 35.1391 respectively, for 4th, 5th and 6th degree polynomial fits. The 6th degree polynomial fit with the lowest residual is showing the best fit for the Country East peak consumption.

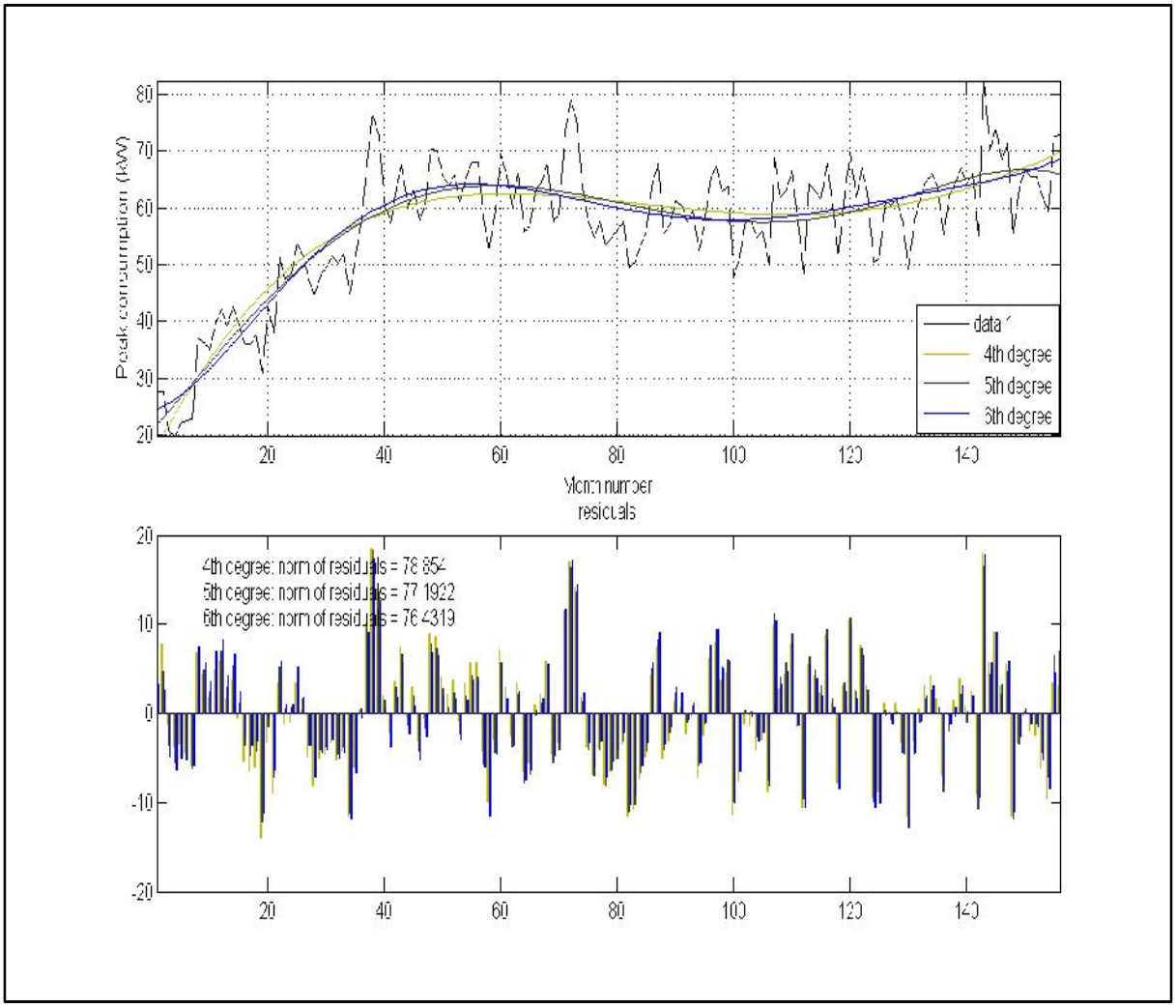


Figure 80: Country East peak detrending.

Figure 81 represents the energy consumption of the Metro South region for thirteen years. The year number fourteen is taken out of the data. It will be used as the test year to evaluate the model. The top pane shows the data set with 4th, 5th and 6th degree polynomial fits. The bottom pane shows the norm of residuals of each fit. All curves seems to be behaving very similar in-terms of capturing the data trend. The lower the norm of the residuals the better the fit will be. In this case the residuals are 233.1321, 229.4652, 228.4735 respectively, for 4th, 5th and 6th degree polynomial fits. The 6th degree polynomial fit with the lowest residual is showing the best fit for the Metro South energy consumption.

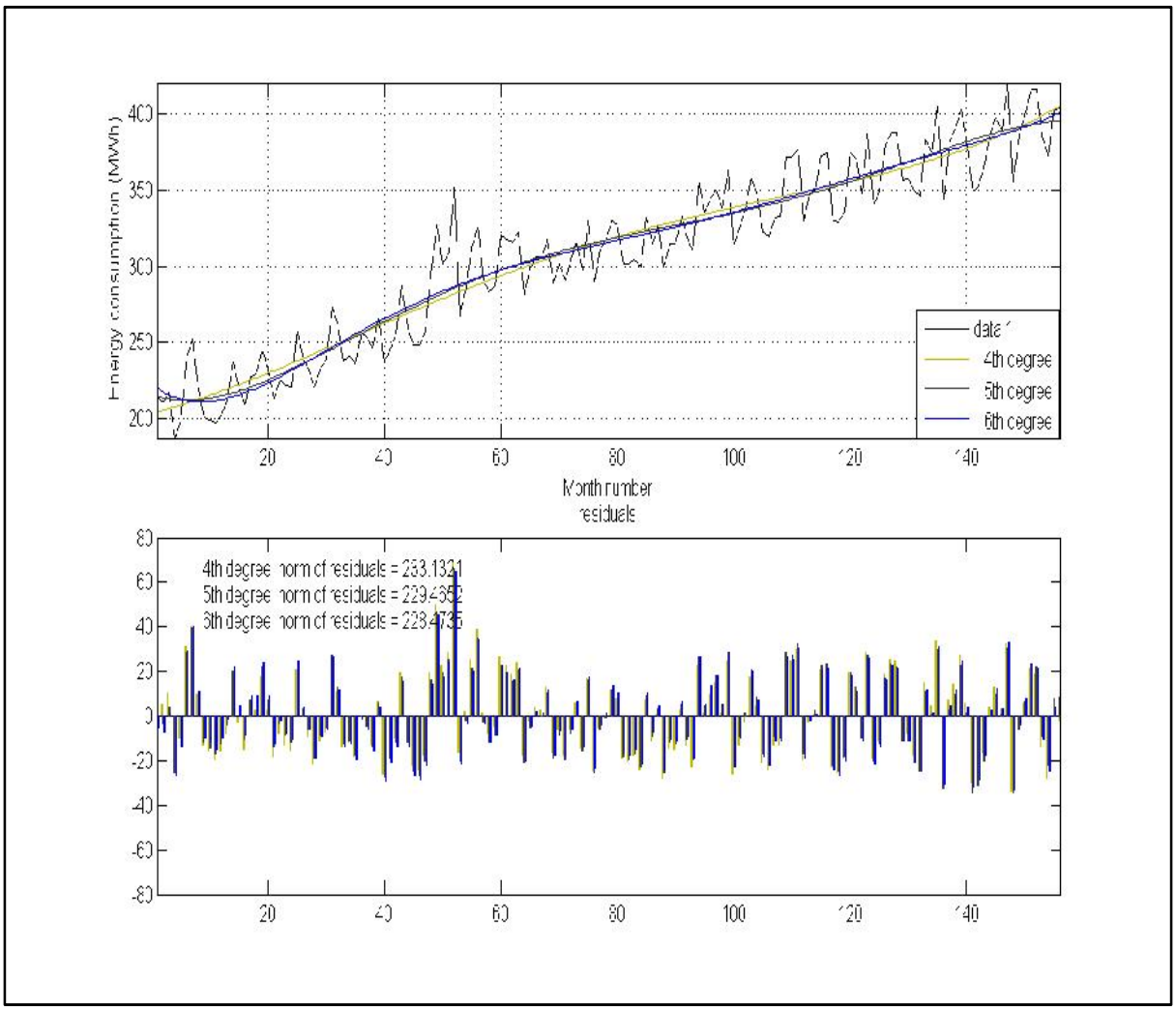


Figure 81: Metro South energy consumption detrending.

Figure 82 represents the peak consumption of the Metro South region for thirteen years. The year number fourteen is taken out of the data. It will be used as the test year to evaluate the model. The top pane shows the data set with 4th, 5th and 6th degree polynomial fits. The bottom pane shows the norm of residuals of each fit. All curves seems to be behaving very similar in-terms of capturing the data trend. The lower the norm of the residuals, the better the fit will be. In this case the residuals are 1067.4553, 1058.5343, 1053.0429 respectively, for 4th, 5th and 6th degree polynomial fits. The 6th degree polynomial fit with the lowest residual is showing the best fit for the Metro South peak consumption.

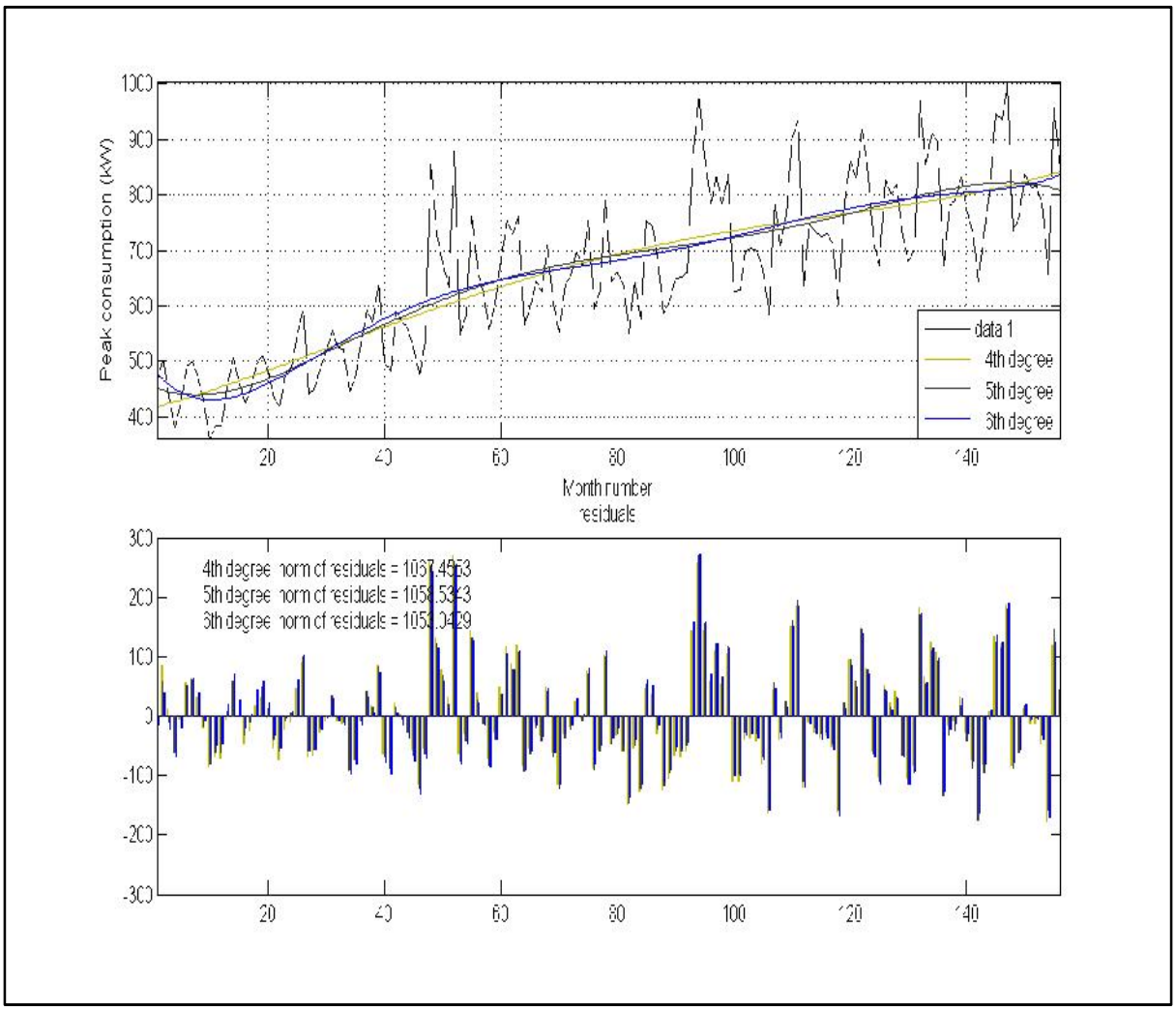


Figure 82: Metro South peak detrending.

Figure 83represents the energy consumption of the Country South region for thirteen years. The year number fourteen is taken out of the data. It will be used as the test year to evaluate the model. The top pane shows the data set with 4th, 5th and 6th degree polynomial fits. The bottom pane shows the norm of residuals of each fit. All curves seems to be behaving very similar in-terms of capturing the data trend. The lower the norm of the residuals, the better the fit will be. In this case the residuals are 233.1321, 229.4652, 228.4735 respectively, for 4th, 5th and 6th degree polynomial fits. The 6th degree polynomial fit with the lowest residual is showing the best fit for Country South energy consumption.

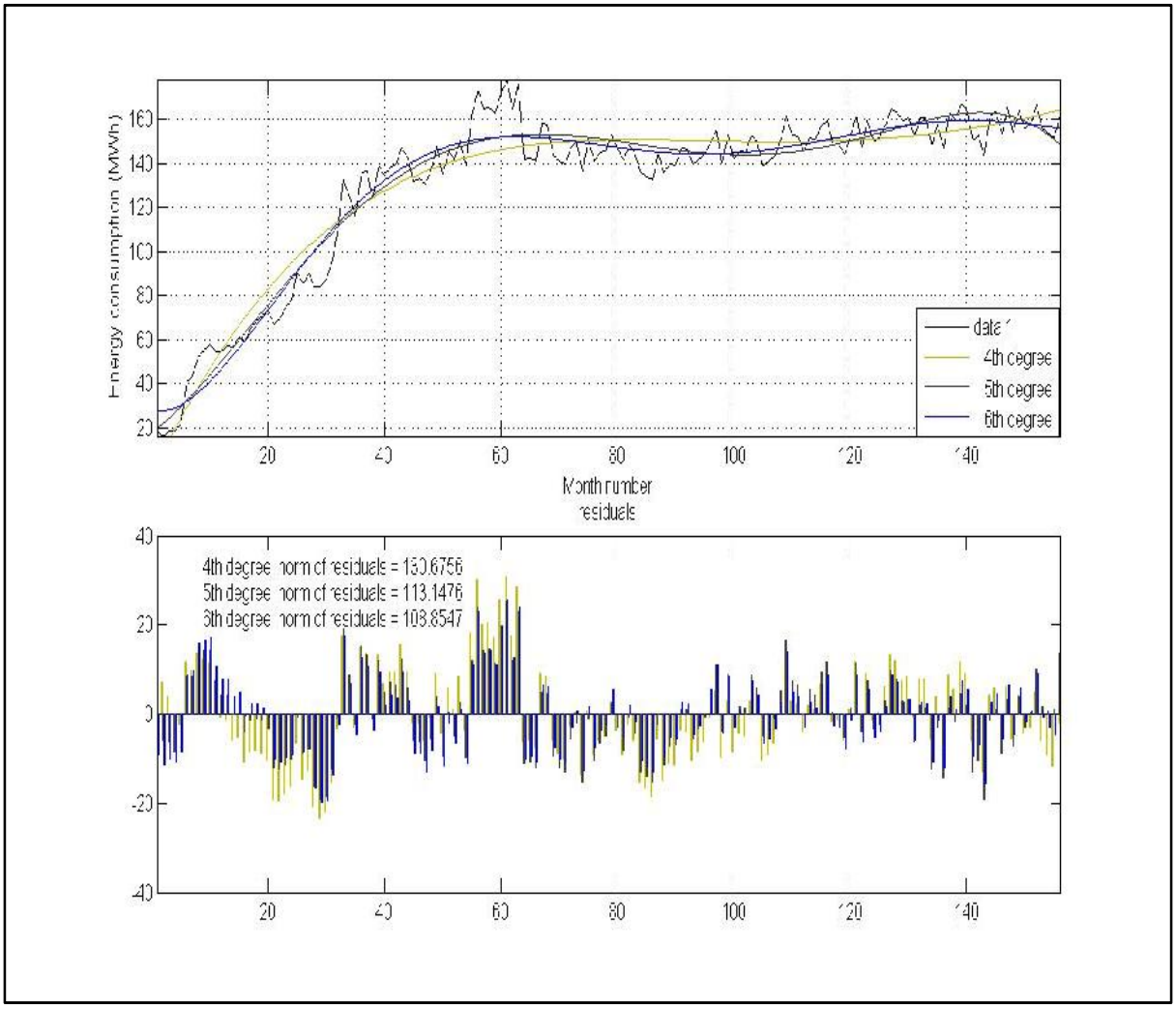


Figure 83: Country South energy consumption detrending.

Figure 84represents the peak consumption of the Country South region for thirteen years. The year number fourteen is taken out of the data. It will be used as the test year to evaluate the model. The top pane shows the data set with 4th, 5th and 6th degree polynomial fits. The bottom pane shows the norm of residuals of each fit. All curves seem to be behaving very similar in-terms of capturing the data trend. The lower the norm of the residuals, the better the fit will be. In this case the residuals are 219.3673, 194.7981, 188.9242 respectively, for 4th, 5th and 6th degree polynomial fits. The 6th degree polynomial fit with the lowest residual is showing the best fit for the Country South peak consumption.

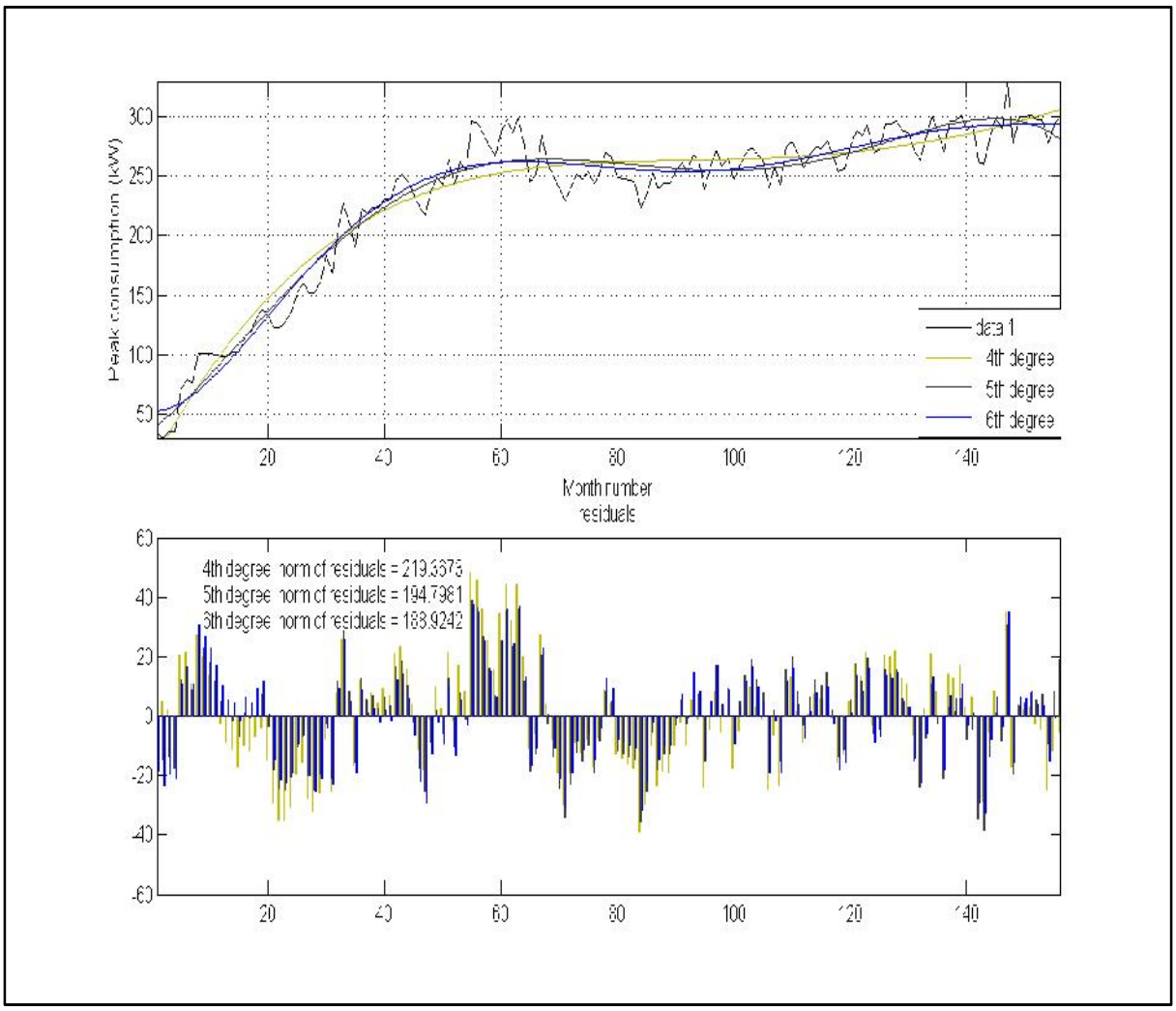


Figure 84: Country South peak consumption detrending.

Figure 85 represents the energy consumption of the Country Goldfields region for thirteen years. The year number fourteen is taken out of the data. It will be used as the test year to evaluate the model. The top pane shows the data set with 4th, 5th and 6th degree polynomial fits. The bottom pane shows the norm of residuals of each fit. All curves seem to be behaving very similar in-terms of capturing the data trend. The lower the norm of the residuals the better the fit will be.In this case the residuals are 103.7916, 99.536, 99.1042 respectively, for 4th, 5th and 6th degree polynomial fits. The 6th degree polynomial fit with the lowest residual is showing the best fit for the Country Goldfields energy consumption.

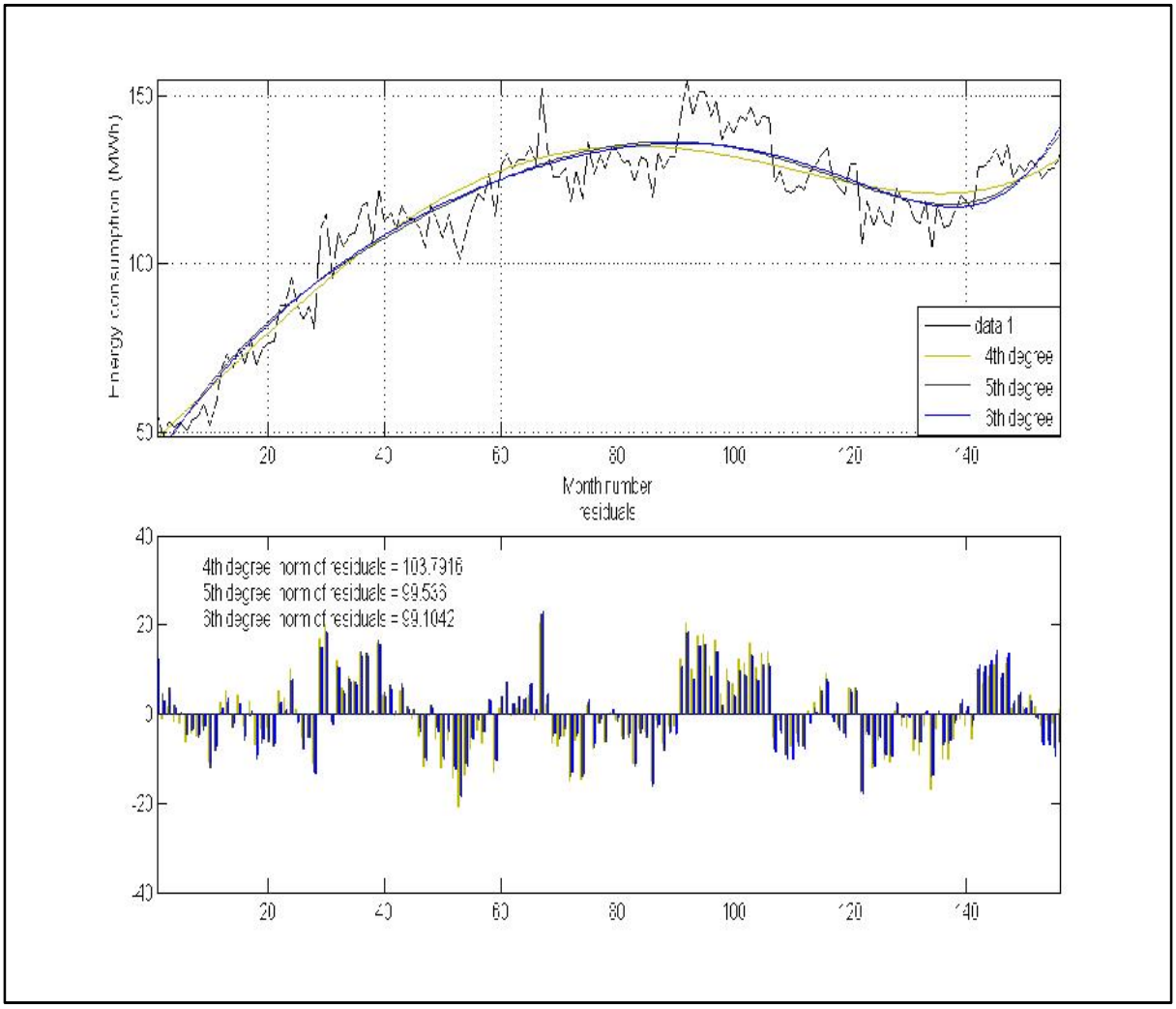


Figure 85: Country Goldfields energy consumption detrending.

Figure 86represents the peak consumption of the Country Goldfields region for thirteen years. The year number fourteen is taken out of the data. It will be used as the test year to evaluate the model. The top pane shows the data set with 4th, 5th and 6th degree polynomial fits. The bottom pane shows the norm of residuals of each fit. All curves seems to be behaving very similar in-terms of capturing the data trend. The lower the norm of the residuals, the better the fit will be. In this case the residuals are 207.1699, 197.0413, 189.7067 respectively, for 4th, 5th and 6th degree polynomial fits. The 6th degree polynomial fit with the lowest residual is showing the best fit for the Country Goldfields peak consumption.

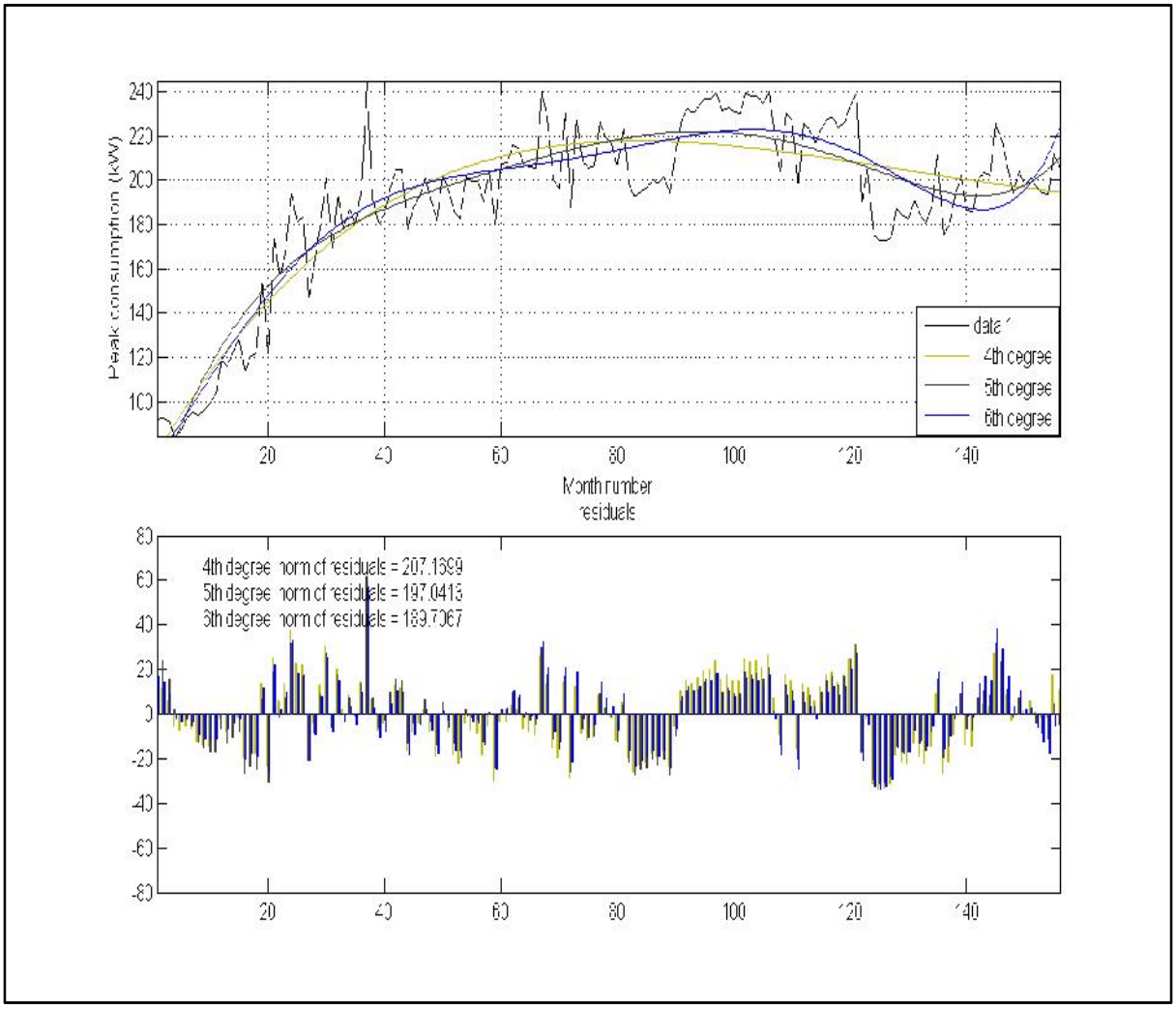


Figure 86: Country Goldfields peak detrending.

Figure 87represents the energy consumption of the Metro North region for thirteen years. The year number fourteen is taken out of the data. It will be used as the test year to evaluate the model. The top pane shows the data set with 4th, 5th and 6th degree polynomial fits. The bottom pane shows the norm of residuals of each fit. All curves seems to be behaving very similar in-terms of capturing the data trend. The lower the norm of the residuals, the better the fit will be. In this case the residuals are 205.4883, 194.7133, 194.13 respectively, for 4th, 5th and 6th degree polynomial fits. The 6th degree polynomial fit with the lowest residual is showing the best fit for the Metro North energy consumption.

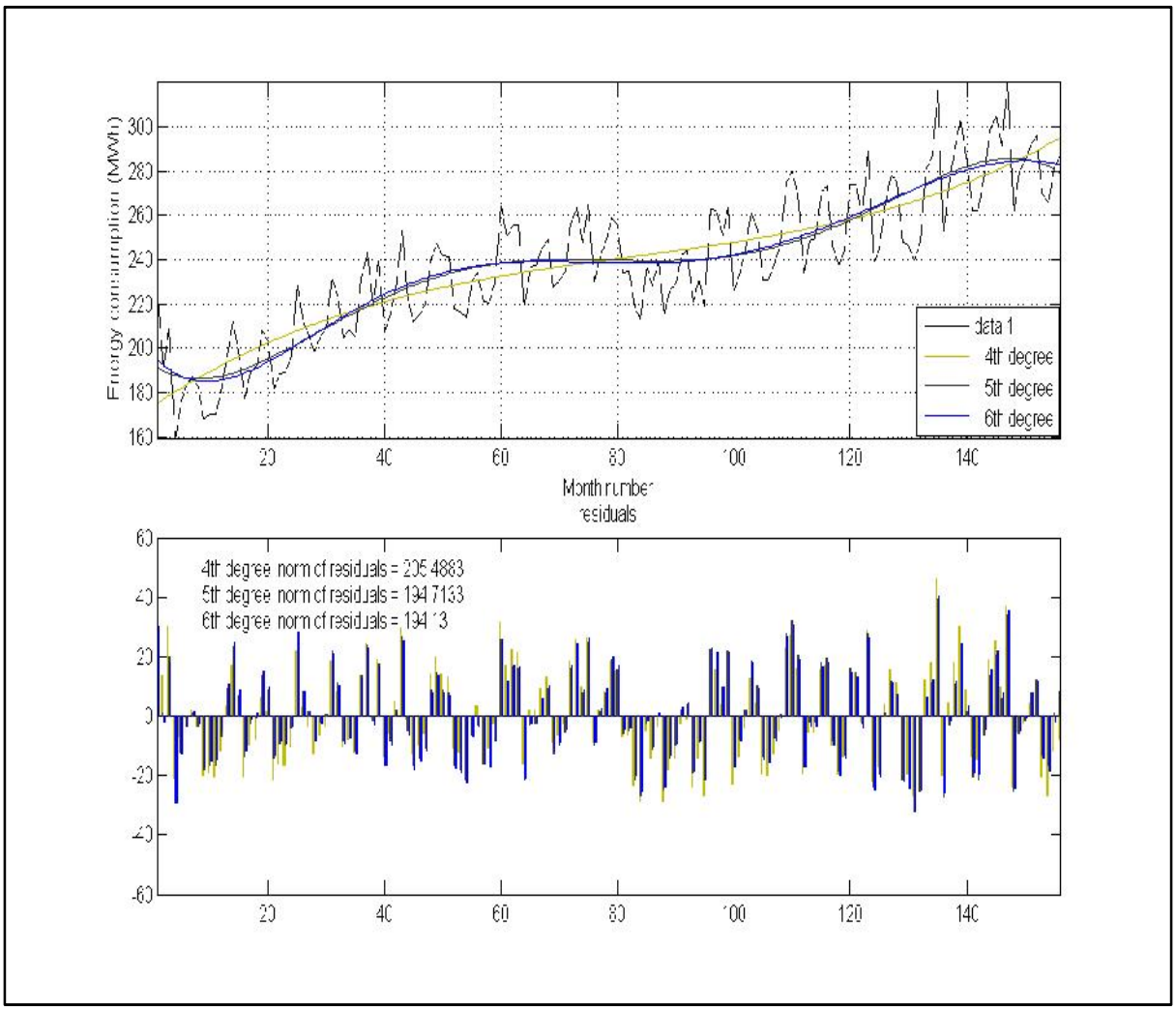


Figure 87: Metro North energy consumption detrending.

Figure 88represents the peak consumption of the Metro North region for thirteen years. The year number fourteen is taken out of the data. It will be used as the test year to evaluate the model. The top pane shows the data set with 4th, 5th and 6th degree polynomial fits. The bottom pane shows the norm of residuals of each fit. All curves seems to be behaving very similar in-terms of capturing the data trend. The lower the norm of the residuals the better the fit will be. In this case the residuals are 829.4341, 821.0249, 813.1431 respectively, for 4th, 5th and 6th degree polynomial fits. The 6th degree polynomial fit with the lowest residual is showing the best fit for the Metro North peak consumption.

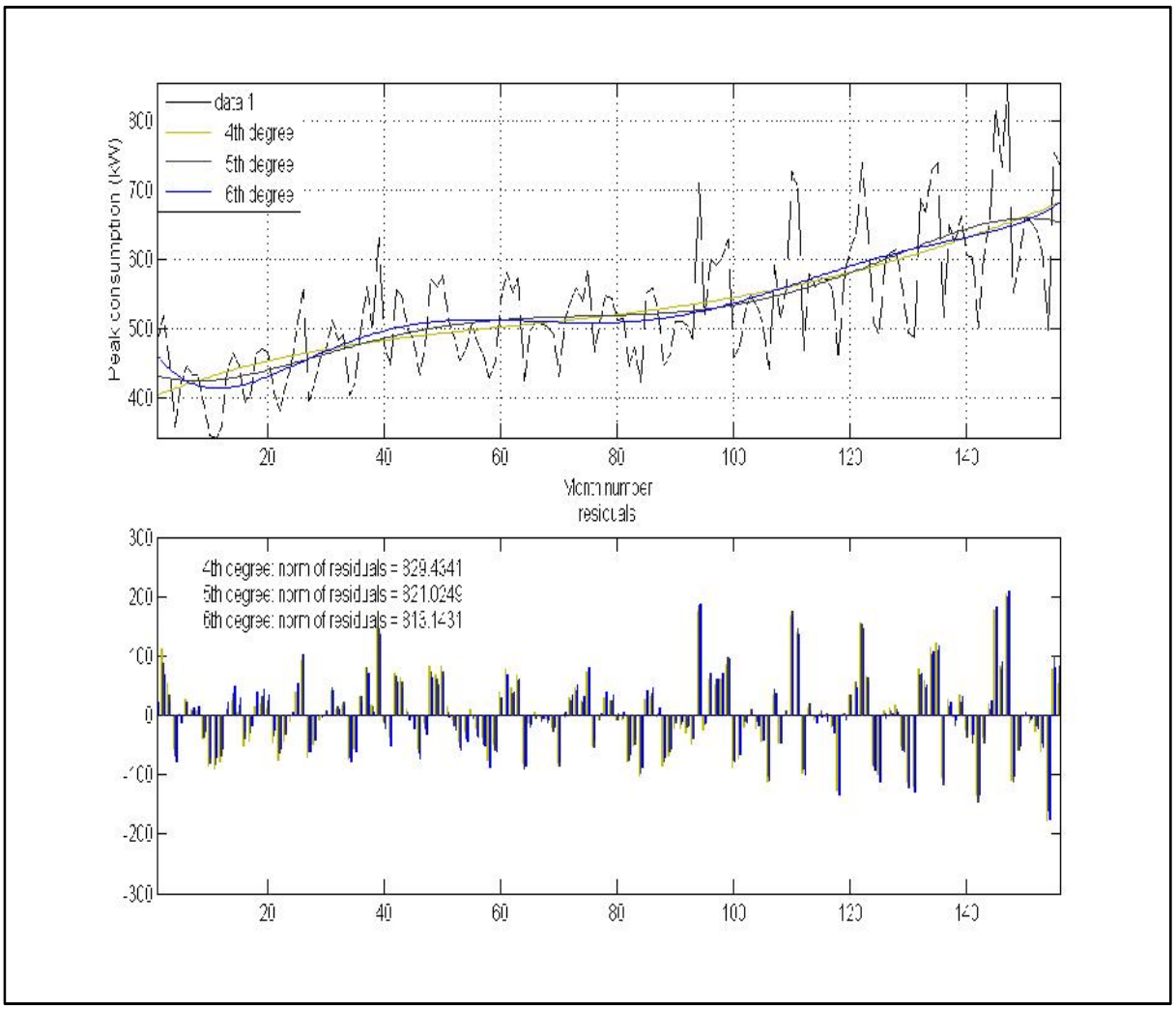


Figure 88: Metro North peak detrending.

Figure 89represents the energy consumption of the Metro East region for thirteen years. The year number fourteen is taken out of the data. It will be used as the test year to evaluate the model. The top pane shows the data set with 4th, 5th and 6th degree polynomial fits. The bottom pane shows the norm of residuals of each fit. All curves seems to be behaving very similar in-terms of capturing the data trend. The lower the norm of the residuals, the better the fit will be. In this case the residuals are 38.3062, 36.648, 36.6411 respectively, for 4th, 5th and 6th degree polynomial fits. The 6th degree polynomial fit with the lowest residual is showing the best fit for Metro East energy consumption.

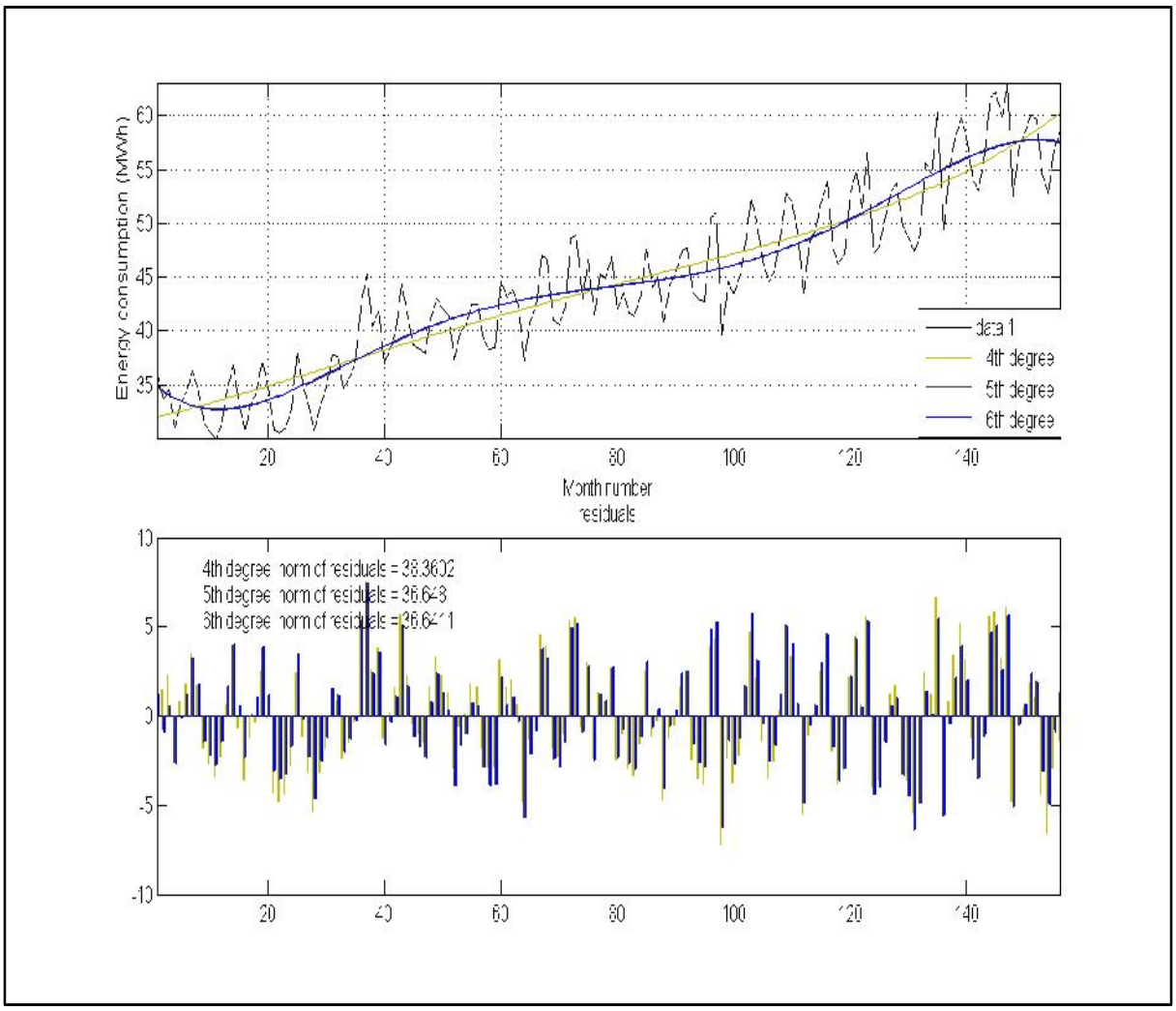


Figure 89: Metro East energy consumption detrending.

Figure 90represents the peak consumption of the Metro East region for thirteen years. The year number fourteen is taken out of the data. It will be used as the test year to evaluate the model. The top pane shows the data set with 4th, 5th and 6th degree polynomial fits. The bottom pane shows the norm of residuals of each fit. All curves seems to be behaving very similar in-terms of capturing the data trend. The lower the norm of the residuals the better the fit will be. In this case the residuals are 130.9652, 128.4241, 127.8772 respectively, for 4th, 5th and 6th degree polynomial fits. The 6th degree polynomial fit with the lowest residual is showing the best fit for Metro East peak consumption.

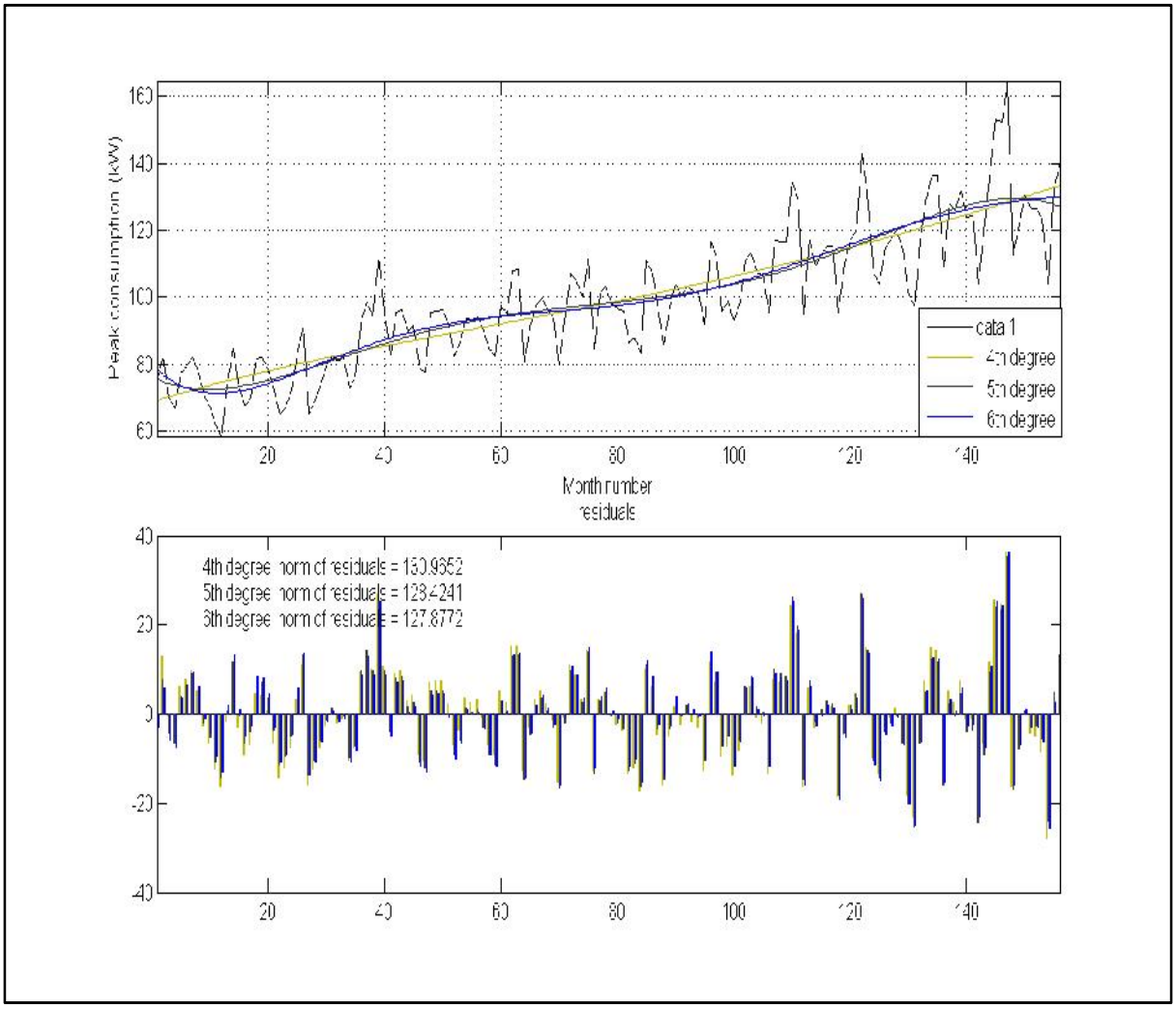


Figure 90: Metro East peak detrending.

It can be seen from the residual norms, the 5th and 6th degree polynomials show a reasonably good level of fit with the smallest value of residual norms. Given that there is a very small change in the norm value from the 5th to 6th degree, and in order to reduce the complexity of the model and to have a consistent detrending strategy for all types of loads, the 5th degree polynomial has been chosen as the detrending method. Refer to Table 14 for details on the fitted polynomials to the sixteen datasets under study.

Table 14: Polynomial fitted to SWIS regions data of peak and energy consumptions.

Load Data	5th degree polynomial
CBD energy (Mwhr)	$Y=-6.1z^5-11z^4+24z^3+8.9z^2-13z+91$
CBD peak (kW)	$Y=-5.4z^5-10z^4+24z^3+3z^2-5z+2.3e^{+2}$
CN energy (Mwhr)	$Y=-0.49z^5-6.4z^4+0.72z^3+14z^2+8.1z+46$

CN peak (kW)	$Y=-2z^5-8.7z^4+4.3z^3+20z^2+12z+80$
CE energy (Mwhr)	$Y=-1.3z^5-7.1z^4+8.4z^3+5.3z^2-7.2z+31$
CE peak (kW)	$Y=-2.2z^5-8.2z^4+14z^3+4.8z^2-9.6z+60$
CS energy (Mwhr)	$Y=-8.9z^5+23z^4+45z^3+5.4z^2-17z+1.5e^{+2}$
CS peak (kW)	$Y=-14z^5-35z^4+71z^3+11z^2-20z+2.6e^{+2}$
CG energy (Mwhr)	$Y=-3.7z^5+0.16z^4+21z^3-17z^2-8.6z+1.3e^{+2}$
CG peak (kW)	$Y=1.6z^5-3.3z^4+7.9z^3-19z^2-2.1z+2.2e^{+2}$
MN energy (Mwhr)	$Y=-9z^5-9.7z^4+36z^3+8z^2-2.8z+2.4e^{+2}$
MN peak (kW)	$Y=-16z^5-2.6z^4+66z^3+16z^2+11+5.2e^{+2}$
ME energy (Mwhr)	$Y=-1.8z^5-0.76z^4+5.9z^3+2.5z^2+5.1z+45$
ME peak (kW)	$Y=-3.4z^5-2.7z^4+10z^3+7z^2+15z+1e^{+2}$
MS energy (Mwhr)	$Y=-5.6z^5-9.5z^4+21z^3-2.8z^2+39z+3.2e^{+2}$
MS peak (kW)	$Y=-19z^5-61z^4+65z^3+26z^2+76z+6.8e^{+2}$

The next step involves deducting the above trends from the available data. The result will be the detrended residuals. The reason for deducting the trends from data is to help with forecast accuracy. As there is not much visible information in residuals graphs and to avoid repeating the figures for all the eight regions, only two regions will be presented here to clarify the steps that has been taken.

Figure 91represents the detrended residuals for the Metro East and Country Goldfields regions. Metro East is predominantly a residential region and Country Goldfield is an industrial one. They are chosen to show that seasonality is present in detrended residuals irrespective of the nature of the load.

The top left pane shows the detrended residuals of the energy consumption of the Country Goldfields region after subtracting the trend polynomial of " Y=-3.7z5+0.16z4 +21z3-17z2-8.6z+1.3e+2 " from the energy consumption data.

The top right pane shows the detrended residuals of the peak consumption of the Country Goldfields region after subtracting the trend polynomial of "Y=1.6z5-3.3z4 +7.9z3-19z2-2.1z+2.2e+2" from the peak consumption data.

The bottom left pane shows the detrended residuals of the energy consumption of the Metro East region after subtracting the trend polynomial of "Y=-1.8z5-0.76z4 +5.9z3+2.5z2+5.1z+45" from the energy consumption data.

The bottom right pane shows the detrended residuals of the peak consumption of the Metro East region after subtracting the trend polynomial of " Y=-3.4z5-2.7z4 +10z3+7z2+15z+1e+2" from the peak consumption data.

As described earlier in the chapter the seasonality happens when the data set repeats itself in constant time intervals. It can be daily, weekly, monthly or yearly. Daily seasonality directly relates to regular daily activities at offices, houses and industrial facilities. Weekly seasonality happens mainly because different sort of activities are being performed during weekends compared to the week days. For industrial consumers who work the same schedule around the week, this type of seasonality won't be present. Monthly seasonality is to do with temperature change over the period of each month. As studied earlier, temperature changes play an important role on the amount of electricity being consumed. And finally, yearly seasonality is mainly to do with public holidays and temperature.

Figure 91 confirms the above statements. As expected, the seasonality of the bottom panes is visible. It can be seen that the data occasionally repeats itself, due to the fast response of residential consumers to the weather fluctuations. Another interesting aspect is that industrial processes are not affected much by changes in weather and electricity consumption is mainly a function of the needs of the industrial processes.

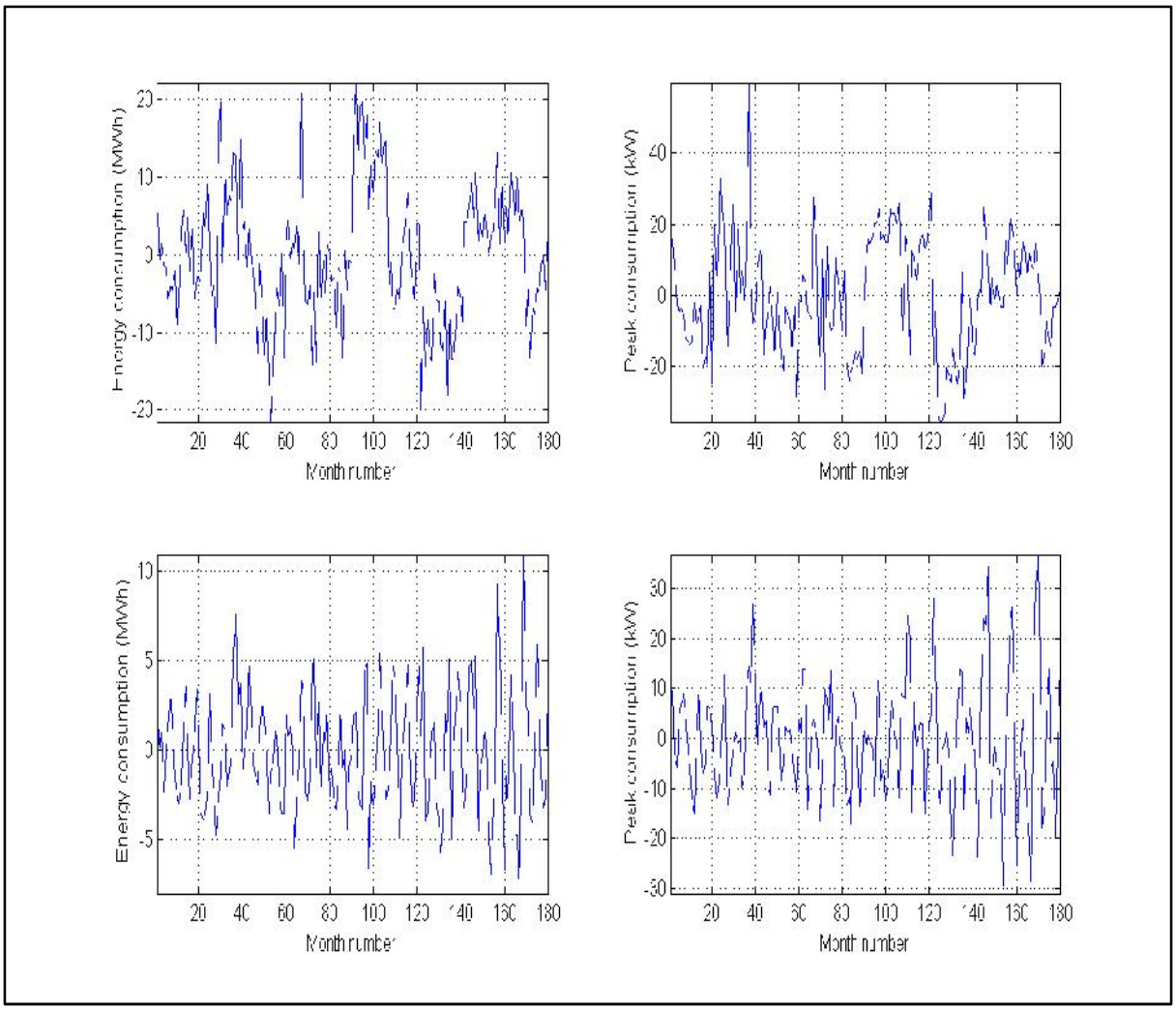


Figure 91: Detrended residuals obtained from deducting trend values from the load data. Top left pane: detrended residuals of the Country Goldfields energy consumption; Top right pane: detrended residuals of the Country Goldfields peak consumption; Bottom left pane: detrended residuals of the Metro East energy consumption; Bottomright pane: detrended residuals of the Metro East peak consumption;

Although weather forecasts of more than one week ahead cannot be accurate, the seasonality of weather is predictable. In other words, in a 24-hour day it is usually warmer at midday and colder at night, and warmer in summer and colder in winter, etc. There is also some seasonality in the behaviours of residential consumers based on the time of the day. For example, residential usage is low at midnight and high during the evening, etc.

If residuals contain visible seasonality, running a training algorithm will likely forecast the seasonality and will not efficiently capture all the small fluctuations of the data. To fix these issues seasonality needs to be removed similar to detrending the datasets.

The Rafal Weron method, introduced in [99], was used to remove the seasonality of the available data. Rafal Weron method which is known as REMST[38] uses the moving average technique as mentioned in section 2.4.3 of [99]. Figure 92shows the residuals after the removal of seasonality. As expected, there is no significant change in industrial residuals. However, in order to be consistent and because some areas have a mixture of industrial, residential and commercial loads, this step was applied to all the regions.

Following the data pre-possessing steps, the detrended deseasonalised residuals can be used as a feed to different training methods.

[38]REMST function can be download for free for MATLAB users from http://fmwww.bc.edu/repec/bocode/r/remst.m

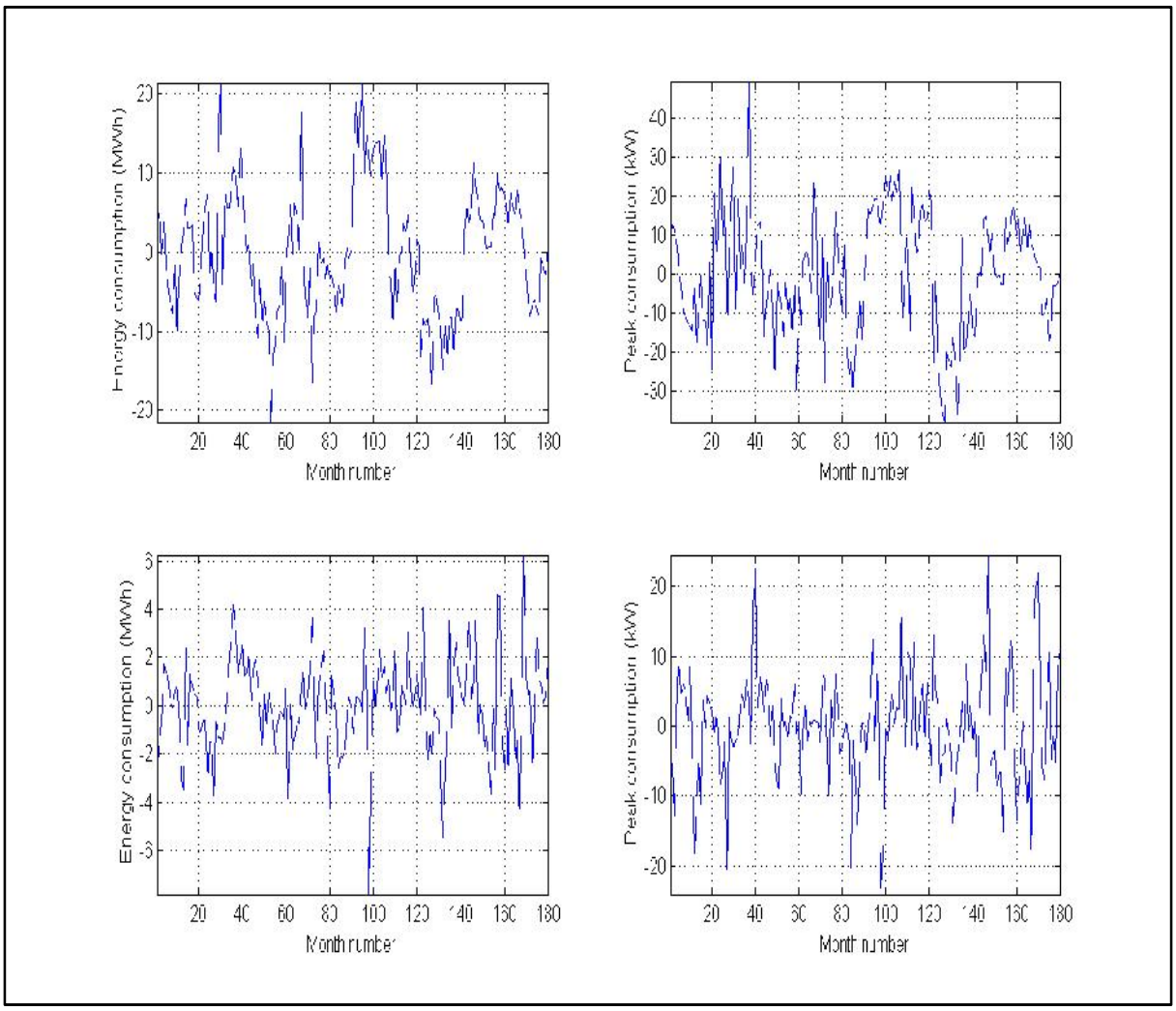


Figure 92: Detrended deseasonalised residuals obtained from removing seasonality from detrended residuals. Top left pane: detrended deseasonalised residuals of the Country Goldfields energy consumption; Top right pane: detrended deseasonalised residuals of the Country Goldfields peak consumption; Bottom left pane: detrended deseasonalised residuals of the Metro East energy consumption; Bottom right pane: detrended deseasonalised residuals of the Metro East peak consumption;

6.4 TRAINING AND RESULTS

The same training algorithms developed in Chapter 5were applied to the residuals of all eight regions of the SWIS. The architecture of the neural network used is feed forward back-propagation with 40 hidden layers and one output layerThe number of layers and type of feed forward network is determined after running a few tests on load data from 1995 to 2008. This configuration had the best regression for both training sets and validation sets. Training performance is set on minimising the mean absolute error.

Decision trees architecture used considers 40 number of regression tree with 30 as the minimum leaf size. Similar to neural network architecture, these numbers are found by trial and error on the data from 1995 to 2008. See codes for more details.

The training method was used to forecast the residuals in the test year and the trend, and seasonality information extracted earlier on was added to the test year. Figure 93shows the complete process in a flowchart. The forecasted value of energy consumption and peak load will be a combination of forecasted residuals, forecasted trend component and forecasted seasonality.

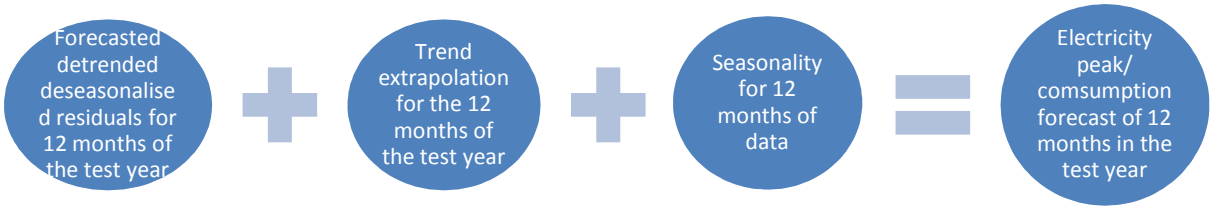


Figure 93: Summary flowchart of medium-term electricity consumption/peak forecast.

The results of the procedure for one year ahead are presented in Figure 94 to Figure 109. Real measured data, neural network forecasted data, and regression tree forecasted data are presented.

Figure 94 represents the medium-term forecast of the energy consumption for the CBD region for the test year of 2010. The black line represents the real data of the test year. The red line shows the forecasted value using Neural networks and the blue one represents the forecasted value using regression trees.Both methods are capable of forecasting the load during the test year. The mean absolute error percentages are 4.47% and 4.24% for Neural networks and regression trees, respectively.

As stated earlier a year ahead forecasts are impossible to generate. Because of that reason much lower accuracies will be acceptable for medium term electricity forecasting. For instance, Ringwood study of Ireland electricity network [126] reports the MAPE value of 12.5%.

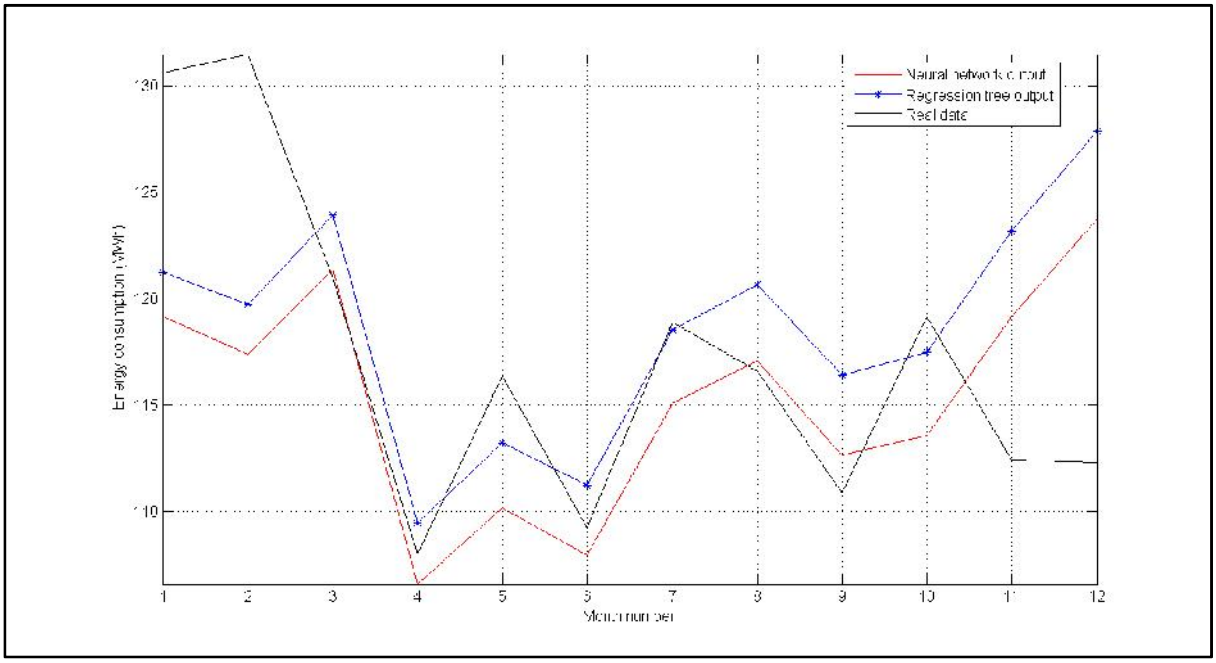


Figure 94: CBD medium-term forecast of energy consumption.

Figure 95 represents the medium-term forecast of the peak consumption for CBD region for the test year of 2010. The black line represents the real data of the test year. The red line shows the forecasted value using Neural networks and the blue one represents the forecasted value using regression trees. Both methods are capable of forecasting the load during the test year. The mean absolute error percentages are 7.94% and 7.41% for Neural networks and regression trees, respectively.

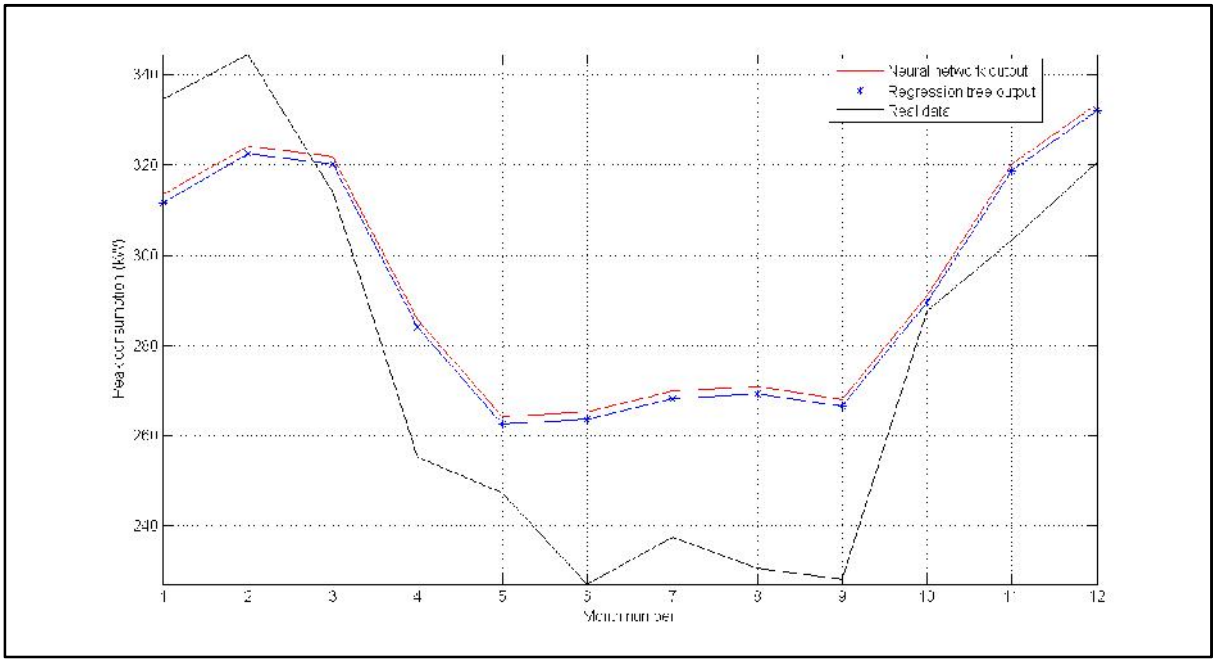


Figure 95: CBD medium-term forecast of peak load.

Figure 96 represents the medium-term forecast of the energy consumption for the Country North region for the test year of 2010. The black line represents the real data of the test year. The red line shows the forecasted value using Neural networks and the blue one represents the forecasted value using regression trees. Both methods are capable of forecasting the load during the test year. The mean absolute error percentages are 3.46% and 3.97% for Neural networks and regression trees respectively.

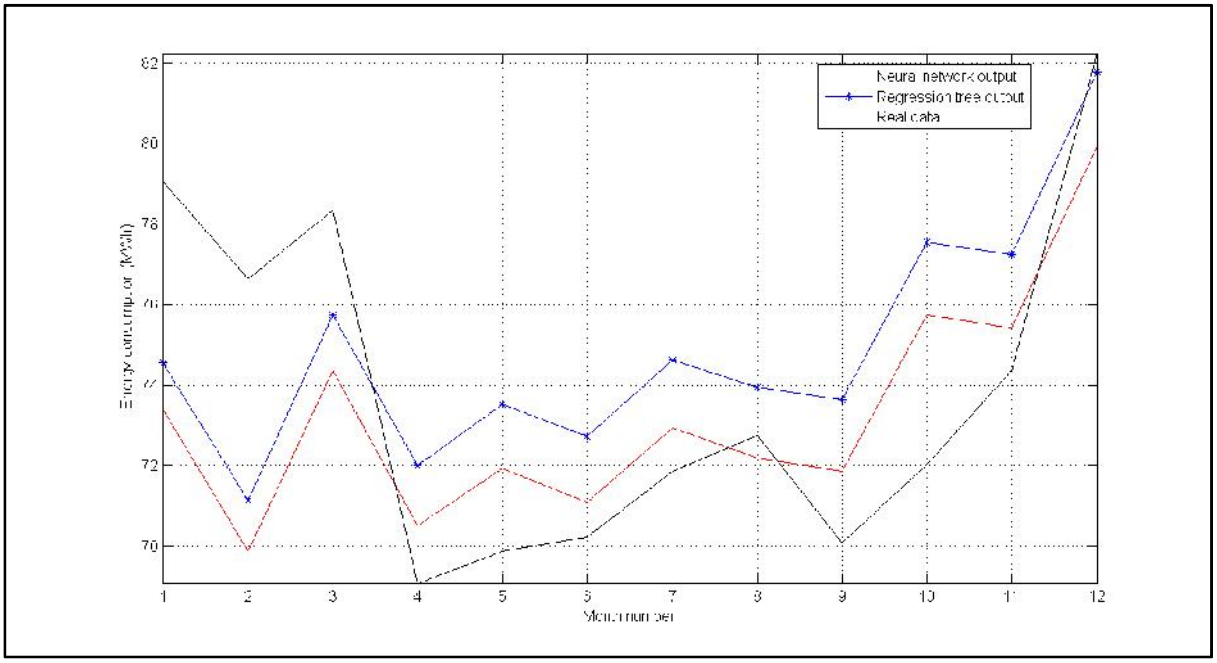


Figure 96: Country North medium-term forecast of energy consumption.

Figure 97 represents the medium-term forecast of the peak consumption for Country North region for the test year of 2010. The black line represents the real data of the test year. The red line shows the forecasted value using Neural networks and the blue one represents the forecasted value using regression trees. Both methods are capable of forecasting the load during the test year. The mean absolute error percentages are 3.45% and 4.23% for Neural networks and regression trees respectively.

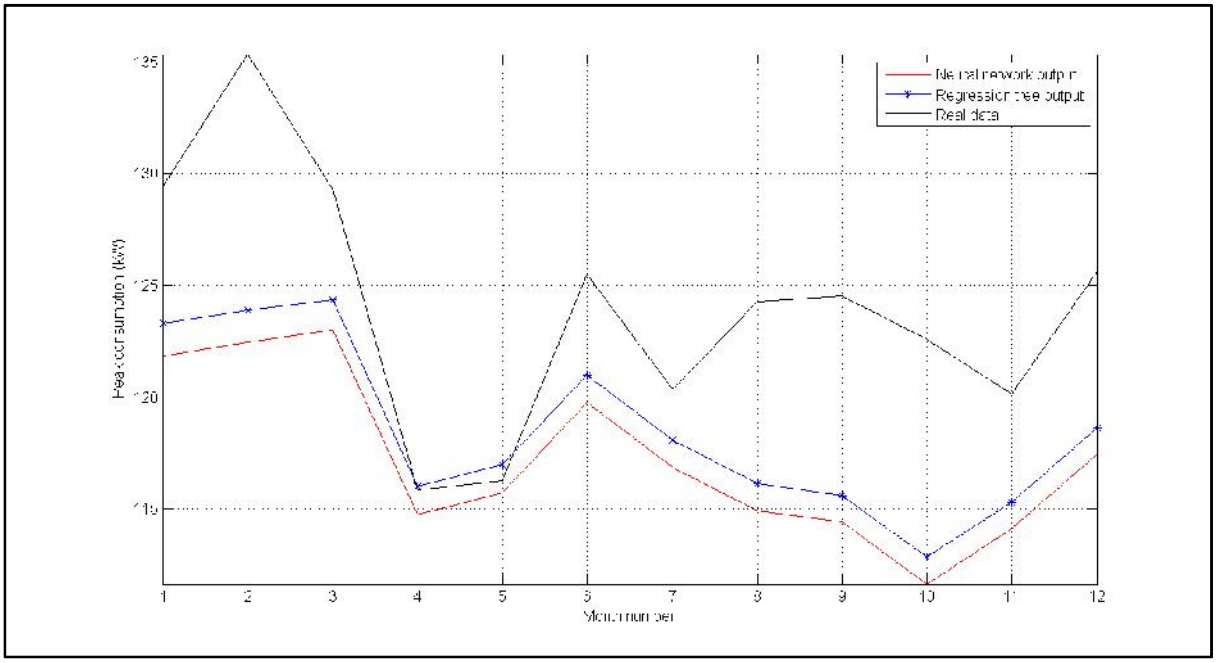


Figure 97: Country North medium-term forecast of peak load.

Figure 98represents the medium-term forecast of the energy consumption for the Country East region for the test year of 2010. The black line represents the real data of the test year. The red line shows the forecasted value using Neural networks and the blue one represents the forecasted value using regression trees. Both methods are capable of forecasting the load during the test year. The mean absolute error percentages are 5.51% and 5.97% for Neural networks and regression trees, respectively.

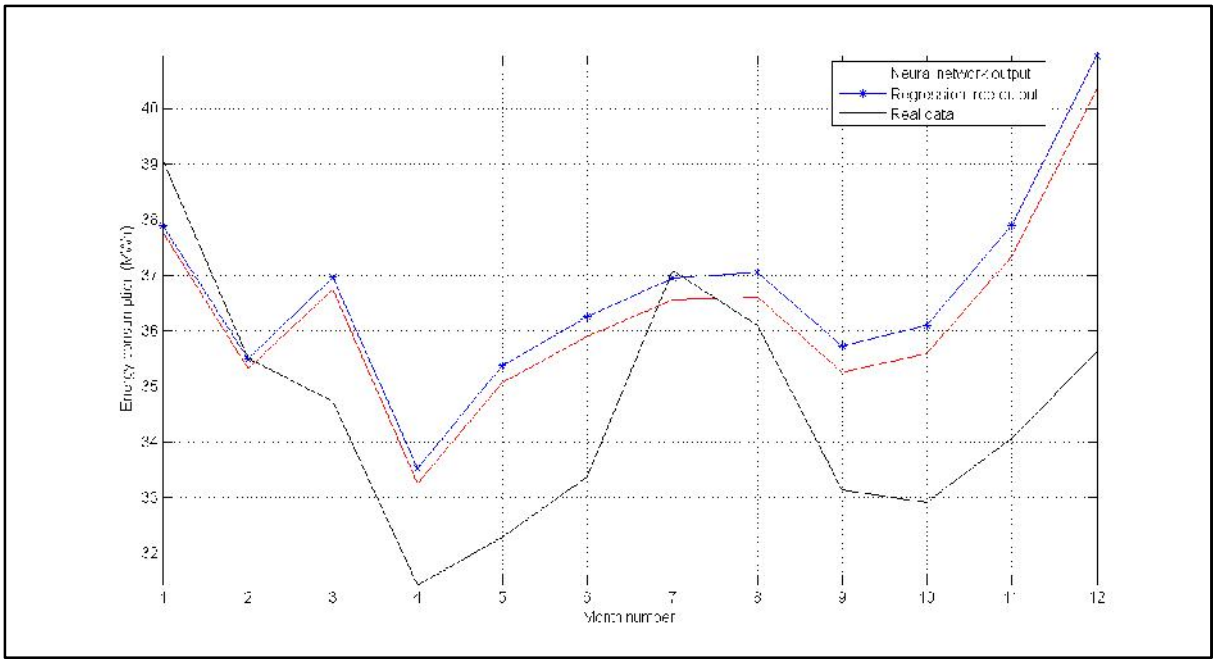


Figure 98: Country East medium-term forecast of energy consumption.

Figure 99 represents the medium-term forecast of the peak consumption for the Country East region for the test year of 2010. The black line represents the real data of the test year. The red line shows the forecasted value using Neural networks and the blue one represents the forecasted value using regression trees. Both methods are capable of forecasting the load during the test year. The mean absolute error percentages are 7.70% and 10.19% for Neural networks and regression trees, respectively.

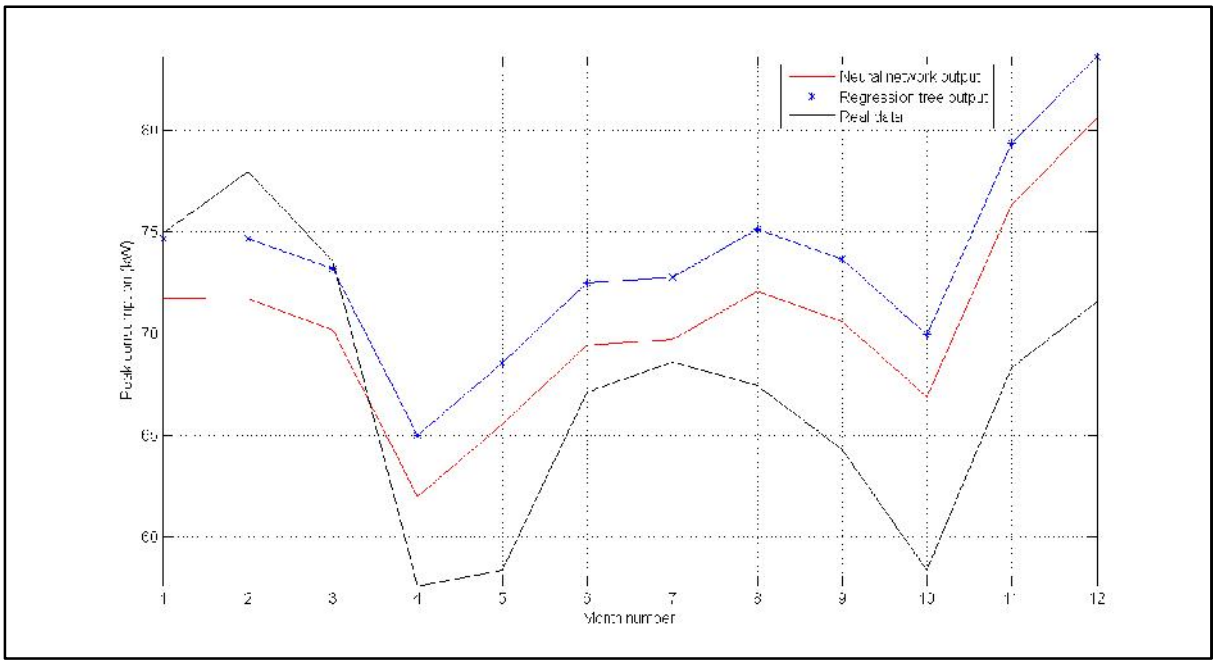


Figure 99: Country East medium-term forecast of peak load.

Figure 100represents the medium-term forecast of the energy consumption for the Country South region for the test year of 2010. The black line represents the real data of the test year. The red line shows the forecasted value using Neural networks and the blue one represents the forecasted value using regression trees. Both methods are capable of forecasting the load during the test year. The mean absolute error percentages are 3.56% and 3.57% for Neural networks and regression trees, respectively.

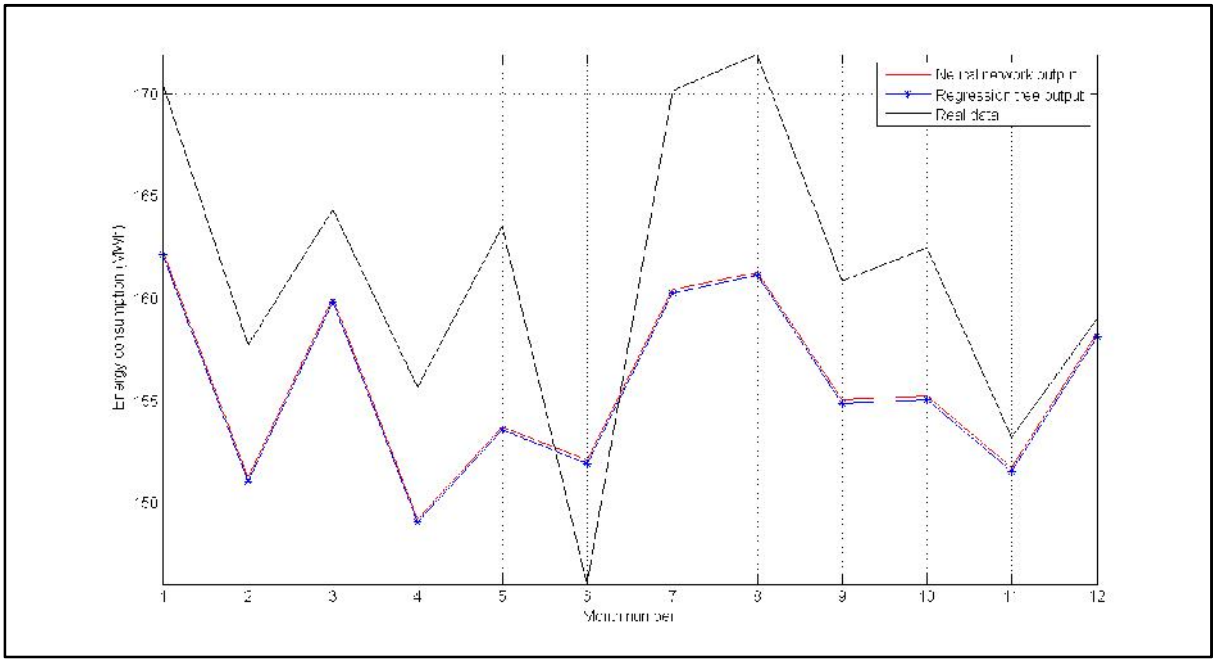


Figure 100: Country South medium-term forecast of energy consumption.

Figure 101 represents the medium-term forecast of the peak consumption for the Country South region for the test year of 2010. The black line represents the real data of the test year. The red line shows the forecasted value using Neural networks and the blue one represents the forecasted value using regression trees. Both methods are capable of forecasting the load during the test year. The mean absolute error percentages are 3.03% and 3.18% for Neural networks and regression trees, respectively.

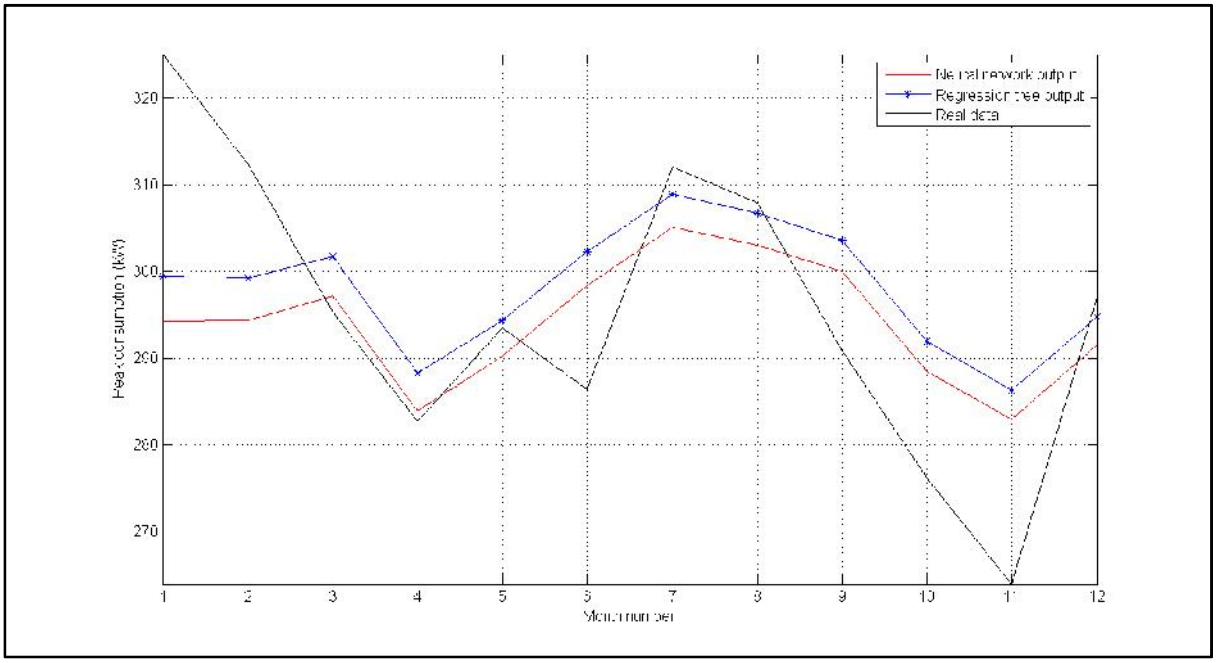


Figure 101: Country South medium-term forecast of peak load.

Figure 102 represents the medium-term forecast of the energy consumption for the Country Goldfields region for the test year of 2010. The black line represents the real data of the test year. The red line shows the forecasted value using Neural networks and the blue one represents the forecasted value using regression trees. Both methods are capable of forecasting the load during the test year. The mean absolute error percentages are 6.53% and 8.94% for Neural networks and regression trees, respectively.

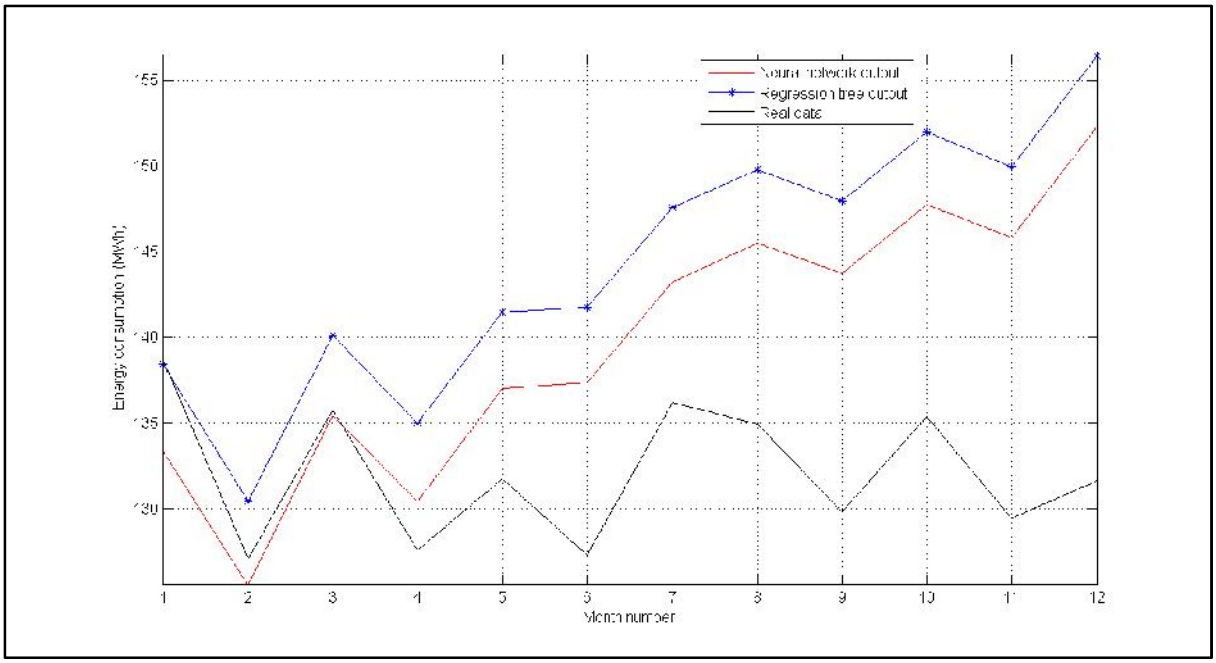


Figure 102: Country Goldfields medium-term forecast of energy consumption.

Figure 103 represents the medium-term forecast of the peak consumption for the Country Goldfields region for the test year of 2010. The black line represents the real data of the test year. The red line shows the forecasted value using Neural networks and the blue one represents the forecasted value using regression trees. Both methods are capable of forecasting the load during the test year. The mean absolute error percentages are 2.61% and 2.85% for Neural networks and regression trees, respectively.

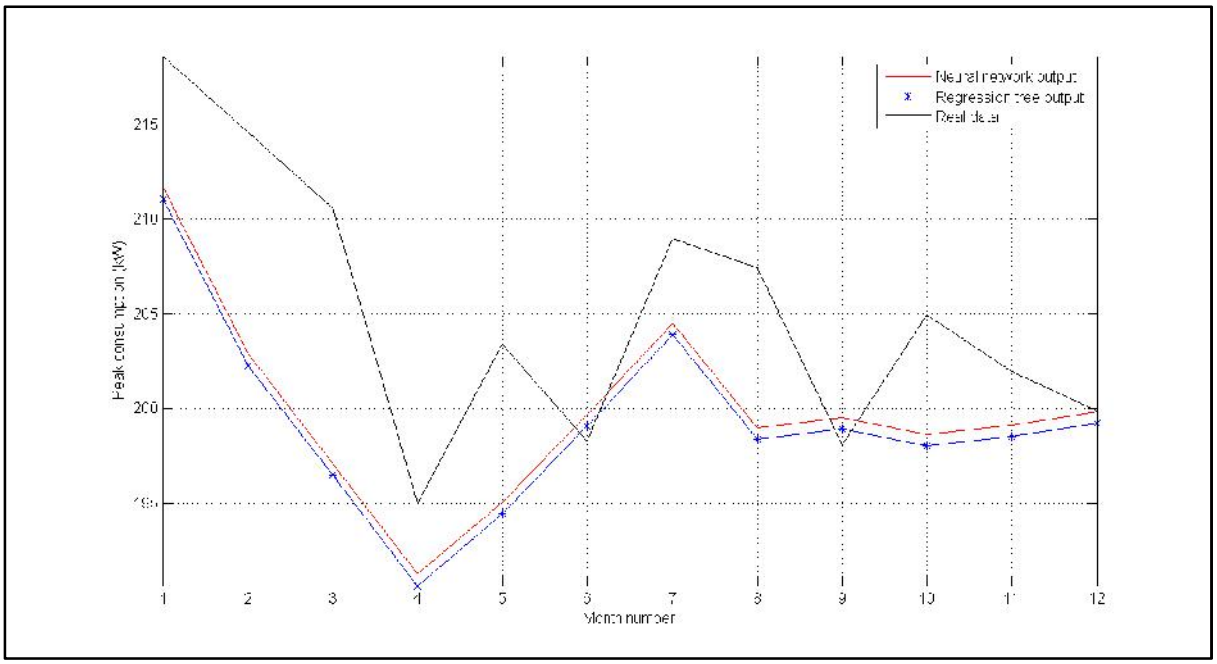


Figure 103: Country Goldfields medium-term forecast of peak load.

Figure 104represents the medium-term forecast of the energy consumption for the Metro North region for the test year of 2010. The black line represents the real data of the test year. The red line shows the forecasted value using Neural networks and the blue one represents the forecasted value using regression trees. Both methods are capable of forecasting the load during the test year. The mean absolute error percentages are 2.67% and 4.51% for Neural networks and regression trees, respectively.

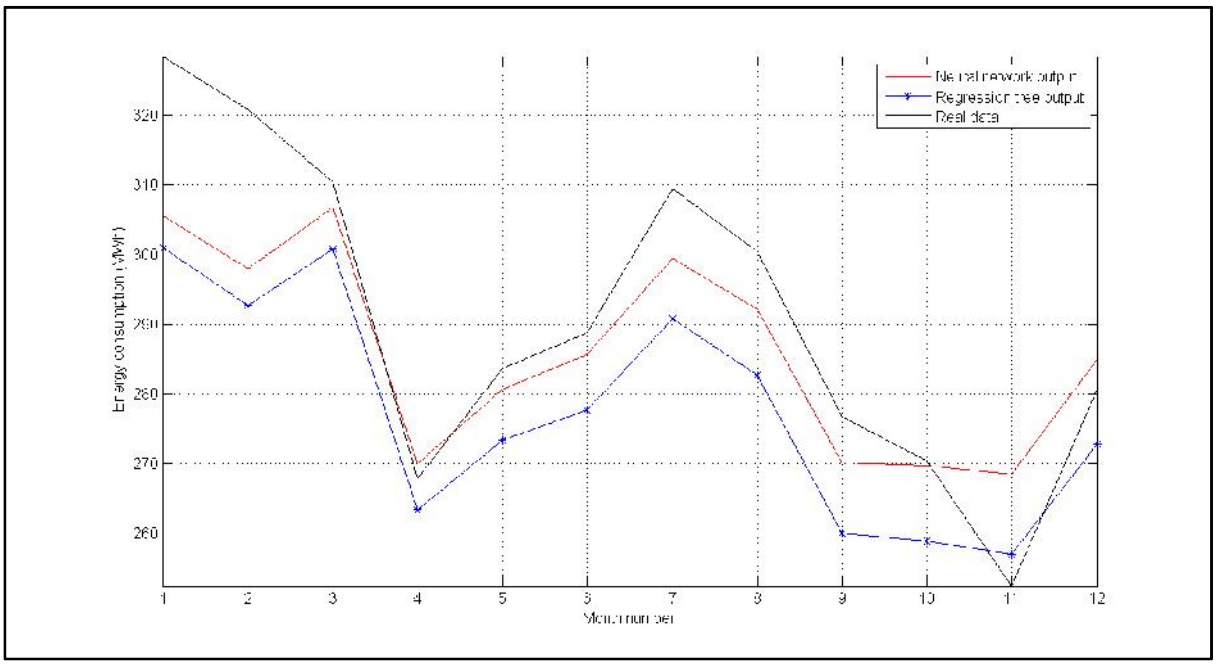


Figure 104: Metro North medium-term forecast of energy consumption.

Figure 105represents the medium-term forecast of the peak consumption for the Metro North region for the test year of 2010. The black line represents the real data of the test year. The red line shows the forecasted value using Neural networks and the blue one represents the forecasted value using regression trees. Both methods are capable of forecasting the load during the test year. The mean absolute error percentages are 8.87% and 8.27% for Neural networks and regression trees, respectively.

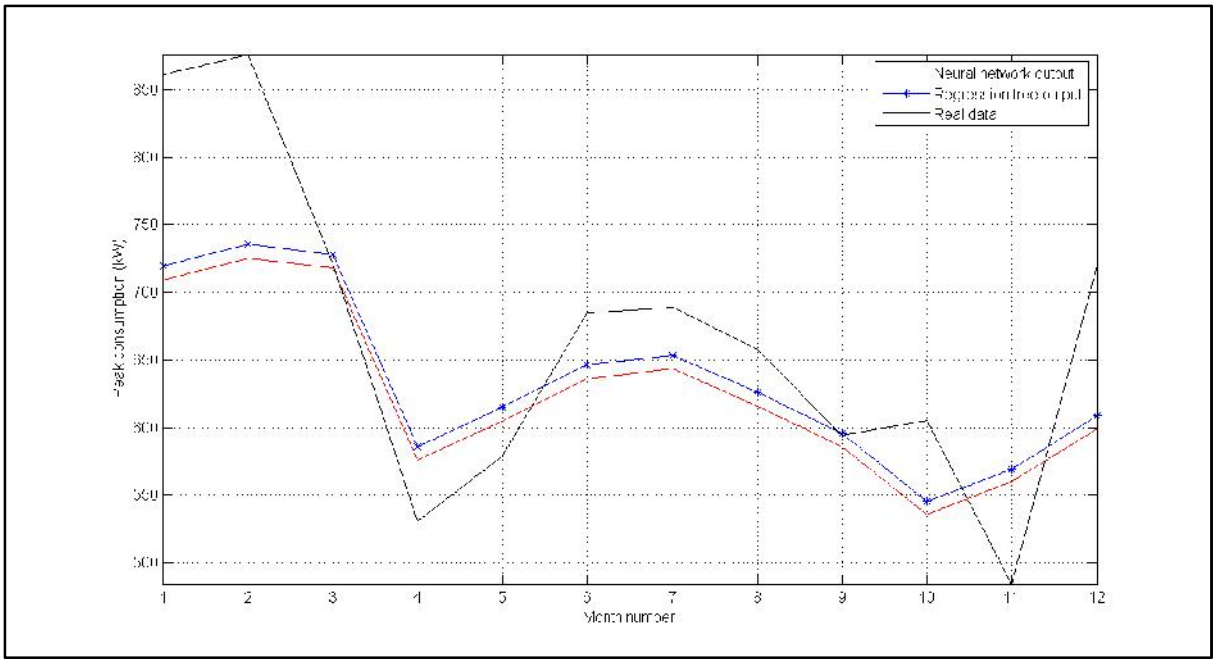


Figure 105: Metro North medium-term forecast of peak load.

Figure 106 represents the medium-term forecast of the energy consumption for the Metro East region for the test year of 2010. The black line represents the real data of the test year. The red line shows the forecasted value using Neural networks and the blue one represents the forecasted value using regression trees. Both methods are capable of forecasting the load during the test year. The mean absolute error percentages are 6.78% and 5.24% for Neural networks and regression trees, respectively.

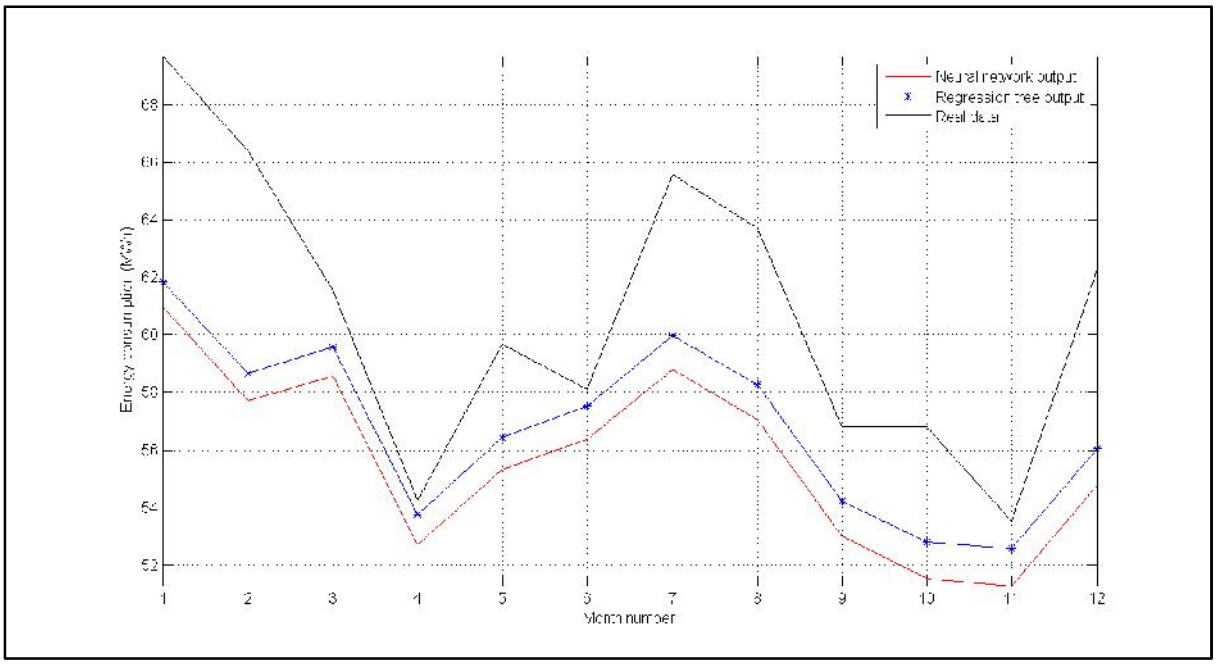


Figure 106: Metro East medium-term forecast of energy consumption.

Figure 107 represents the medium-term forecast of the peak consumption for the Metro East region for the test year of 2010. The black line represents the real data of the test year. The red line shows the forecasted value using Neural networks and the blue one represents the forecasted value using regression trees. Both methods are capable of forecasting the load during the test year. The mean absolute error percentages are 14.3% and 8.26% for Neural networks and regression trees, respectively.

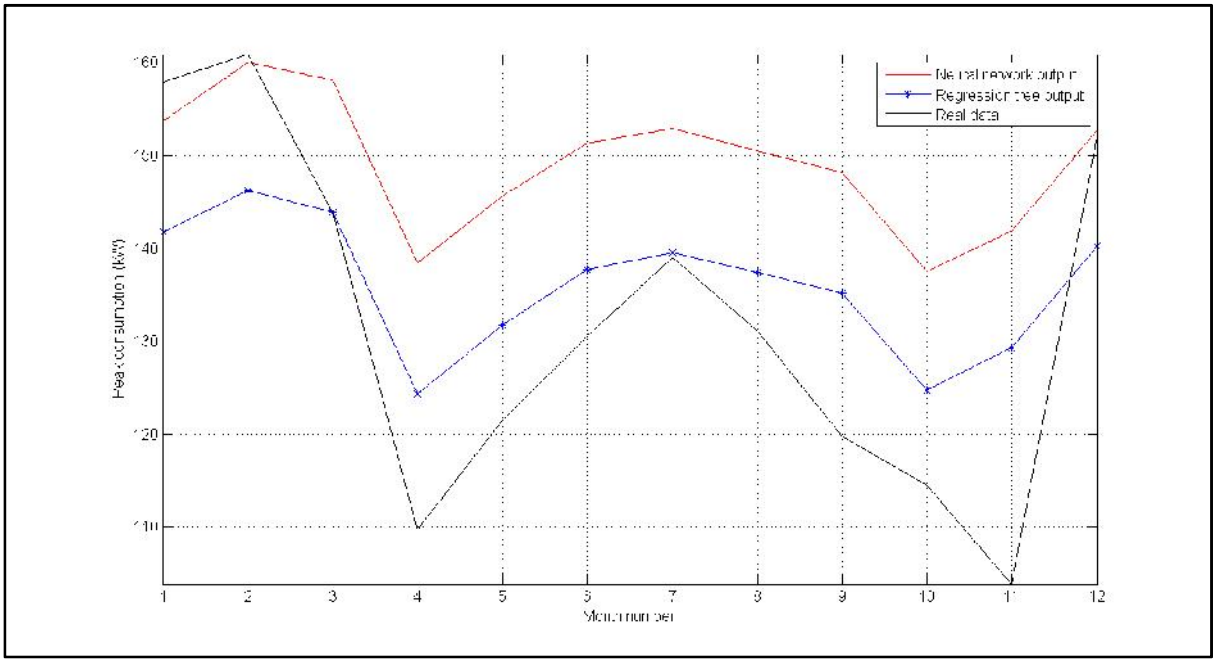


Figure 107: Metro East medium-term forecast of peak load.

Figure 108 represents the medium-term forecast of the energy consumption for the Metro South region for the test year of 2010. The black line represents the real data of the test year. The red line shows the forecasted value using Neural networks and the blue one represents the forecasted value using regression trees. Both methods are capable of forecasting the load during the test year. The mean absolute error percentages are 5.09% and 2.93% for Neural networks and regression trees, respectively.

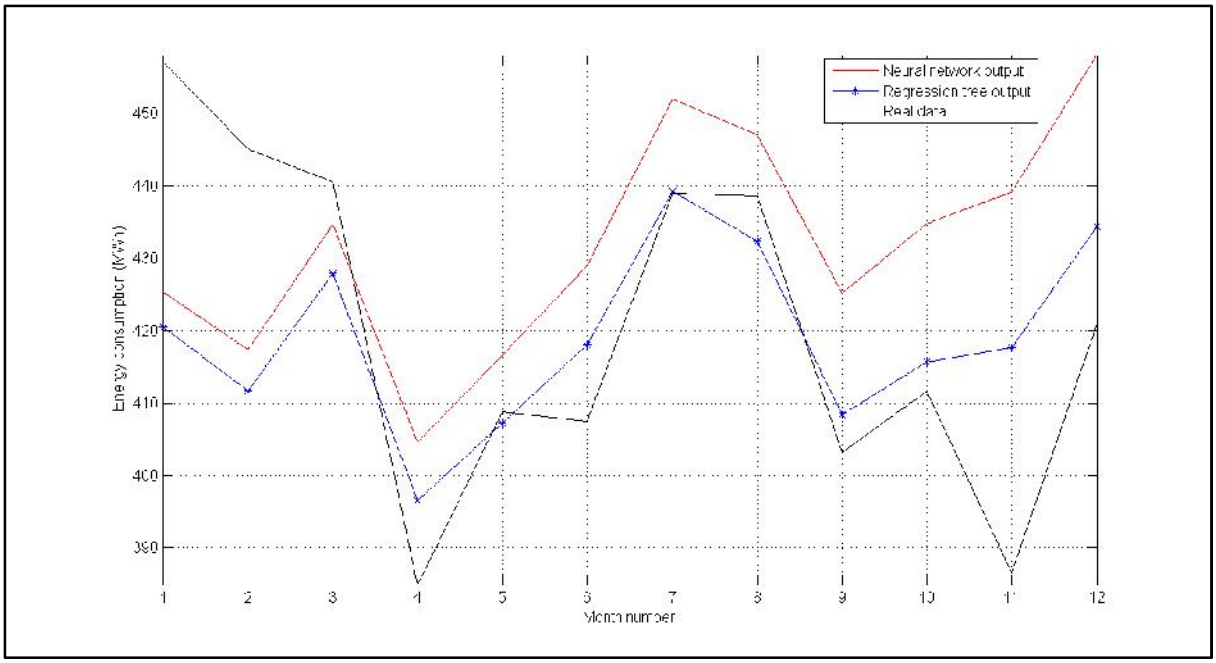


Figure 108: Metro South medium-term forecast of energy consumption.

Figure 109 represents the medium-term forecast of the peak consumption for the Metro South region for the test year of 2010. The black line represents the real data of the test year. The red line shows the forecasted value using Neural networks and the blue one represents the forecasted value using regression trees. Both methods are capable of forecasting the load during the test year. The mean absolute error percentages are 4.91% and 6.56% for Neural networks and regression trees, respectively.

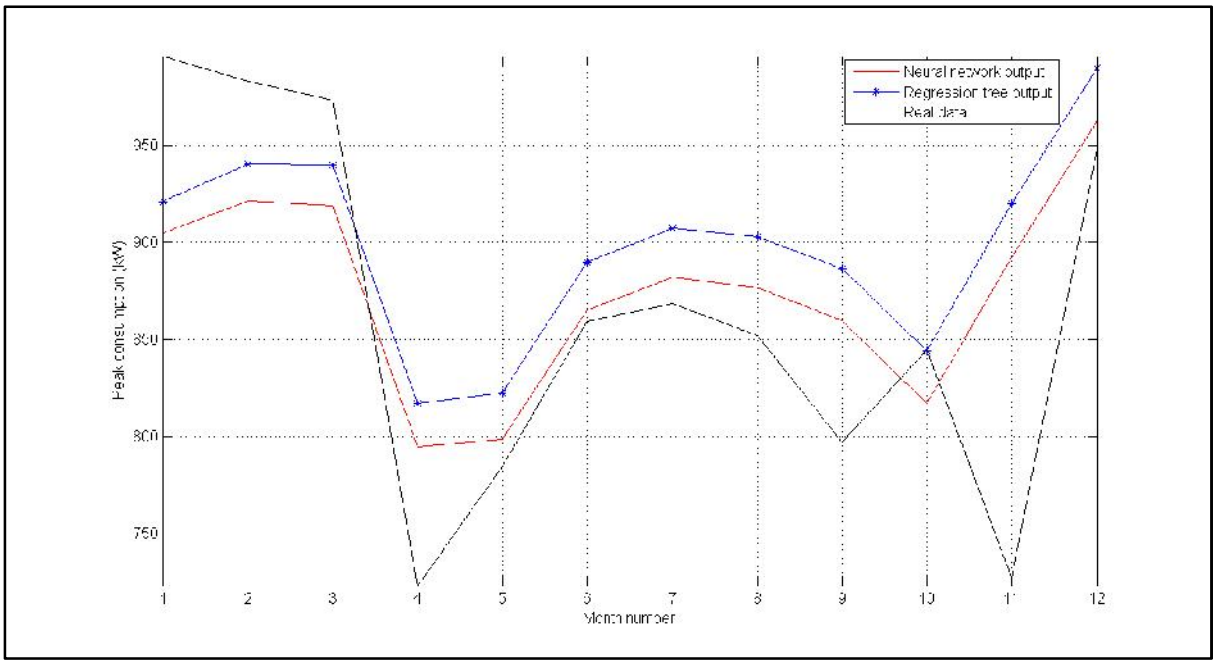


Figure 109: Metro South medium-term forecast of peak load.

6.5 CONCLUSION

Table 15 presents the MAPE of neural network and regression trees for a year ahead forecasts. The table includes forecasts for both peak demand and energy consumption. The MAPE errors are very different for different regions. They start from as little as 2.61% up to 14.3%. with only one exception, all the other sixteen MAPE values are below the 12.5% reported by Ringwood on his studies of Ireland electricity network [126].

The second observation is that the energy consumption forecasts are generally more accurate than the peak demand forecasts. This was anticipated because of the more random behaviour of peak consumption[39]. The only exception is the Country Goldfields region where peak demand forecast is more accurate than the energy consumption forecast. Given that the electricity consumption of the region is mainly because of the mines located in the area, the discrepancy may be due to a scheduled maintenance that lasted less than a month or a reduction in the

[39]From mathematical point of view, the energy consumption is the integral of instantaneous load consumed as a function of time. This makes the energy consumption graph to be always smoother than the instantaneous load graph.

number of working shifts in one of the major mines. Unfortunately, this information is not available and cannot be discussed in more detail.

The other interesting observation is the big change in MAPE values from as little as 2.61% to 14.3% for different case studies. Bear in mind that the method used is the same and the difference is the case study used. It can be concluded that accuracy of a method used on a case study cannot guarantee the similar sort of accuracies when being applied on a different case study. For example studies used on Ireland electricity market with high level of accuracies can perform poorly on Australian electricity market and vice versa. Therefore a tailor made forecasting method is required for each electricity system under study.

Table 15:A year ahead MAPE of peak demand and energy consumption for all the eight regions of SWIS

Region	NN Based MTLF MAPE	RT Based MTLF MAPE
CBD Energy Consumption	4.47%	4.24%
CBD Peak Load	7.94%	7.41%
Metro North Energy Consumption	2.67%	4.51%
Metro North Peak Load	8.87%	8.27%
Metro East Energy Consumption	6.78%	5.24%
Metro East Peak Load	14.3%	8.26%
Metro South Energy Consumption	5.09%	2.93%
Metro South Peak Load	4.91%	6.56%
Country North Energy Consumption	3.46%	3.97%
Country North Peak Load	3.45%	4.23%
Country East Energy Consumption	5.51%	5.97%
Country East Peak Load	7.70%	10.19%
Country South Energy Consumption	3.56%	3.57%
Country South Peak Load	3.03%	3.18%
Country Goldfields Energy Consumption	6.53%	8.94%
Country Goldfields Peak Load	2.61%	2.85%

For industrial regions regression trees performs better than neural network based method. These regions are Country North, Country East, Country South and Country Goldfields. The same applies to CBD region where commercial load dominates. For some residential areas neural networks behave better. This is because of higher nonlinearity of residential load. To summarize and because all the regression tree based MAPE values are less than 10% it can be concluded that regression tree based method is more reliable for medium-term forecasts.

The proposed methodology provides a very good accuracy to medium-term forecasts of one year ahead. The results are reliable, and in most cases, are within10% margin of the real data. There is only one exception among the case studies, namely, the peak demand forecast of the Metro East region using neural networks. The forecast of the same variable using regression trees has an MAPE of 8.26%.

Chapter

7 SUMMARY AND FUTURE RECOMMENDATIONS

7.1 THE MOTIVATION OF THE WORK

The main motivation of this research is to help reduce the GHG emissions of the electricity sector, and counteract the effects on nature and people. Traditional methods of power planning are not optimised to achieve this, and consider Capex and Opex reduction as their main objectives [38]. Minimising GHG emissions is now an additional objective of power planning [1], [2]. One way of achieving this is by optimising the distance of generators to the loads to reduce the transmission losses, and also by harnessing the available regional sources of renewable energies and increasing their integration in the network [39]. Efficient load forecasting methods capable of describing the regional behaviours of the electricity consumption are developed in this research, and can provide priceless input to electricity planners. Such forecasting methods, known as spatial forecasting [40], can be used to extract short-term and medium-term information of the electricity consumption of different regions. This work also provides tools for making decisions about the most accurate way of pre-processing consumption data and choosing the most efficient forecasting procedure.

7.2 MAJOR CONTRIBUTION OF THE WORK

Major contribution of the work can be summarised as below.

- The topic of the study, i.e. spatial load forecasting and the potential of using it in efficient power planning, is relatively a new topic in the electricity market literature. Moreover, many of the known spatial load forecasting methods have not yet been widely used because of the size, variety, and availability of the data required. The methodology proposed in this study can successfully be applied to spatial forecasting.

- While conventional methods are useful for short-term predictions with acceptable accuracy, they fail when medium-to-long term load forecasting is dealt with. The methodology conceived and implemented in this thesis is significantly better than those known as state-of-the-art and can give very satisfactory results for medium-term predictions.

- The load analysis criterion, particularly using Q-Q (Quantile vs. Quantile) plots is a unique and original finding of this work. While Q-Q plots are largely used in traditional statistics to compare two samples of data, it has never been applied before for electricity load forecasting purposes. Based on its definition and use, an electricity planner can understand which part of the load is the dominating factor (i.e. whether it is residential, commercial or industrial). And then, based on this, he/she can decide how to go ahead with choosing the most effective forecasting method. Based on this, the thesis provides a very useful criterion for decision making in the energy market.

- One of the major findings of the thesis is that there is no one optimum way of forecasting electricity load in different scenarios. The results presented in the thesis have shown that a method that can accurately forecast the load on a system (3% error for a year ahead) can perform completely different in forecasting another system (observed errors of around 14%).

This study demonstrates that a method which is claimed to have a given accuracy can be considerably inaccurate when applied on a different case study.

- Using an ambient temperature-based criterion (i.e. the average maximum temperature of the month) to choose the correct forecasting method is another major finding of the study. In fact, the author has demonstrated that for a temperature sensitive load, different forecasting methods should be used and then combined to get the most accurate result.

7.3 THE STEPS OF THE PROJECT

Chapter 1 talks about emissions of greenhouse gases and their adverse effect on the nature. It introduces electricity sector as one of the major contributors of human made GHG emissions. It then describes the components of electrical power network and the planning of it. Finally the chapter concludes that an efficient spatial load forecasting method is required to help with spatial planning of power networks. The spatial planning can include more regional components like proximity of generation components to consumers, or the levels of harnessed renewable energy in each area. In such an approach, GHG reduction can be also considered along with Capex and Opex minimisation to plan the future of power networks.

Chapter 2 provides definitions on power network components and the load forecasting methods. It starts with definition of power systems and explanation on how electrical energy is superior to all other forms of energy from end user point of view. Electricity generation systems and the sources of energy to produce electricity are described next. Typical generation unit sizes in MW, continuity of the supply, and also its predictability are summarised in a table at the end of this section. Then, transmission lines and distribution systems are described, as other component of electrical power networks. Importance of having an accurate forecast of electricity demand and the common ways to do it are presented next. At the end of this chapter the deficiencies of current forecasting methods are highlighted and one major goals is defined for this work. It is to overcome the deficiencies of individual forecasting methods by combining them and using them only where it performs efficient. It also mentions that the work is going to closely look at the behaviour of input data to the forecasting method to seek better methods for preparing them.

Chapter 3 describes SWIS as the case study for this work. The reasons for selecting SWIS as the case study are mentioned, followed by a quick history of it and how it has been expanded over the last hundred years. To be able to complete spatial forecasting, the area under study needs to be divided into regions. SWIS is then divided into eight regions for this purpose. A visual presentation of the eight regions on the map is presented at the end of this chapter for more clarity.

Chapter 4 performs a short forecasting method on one of the SWIS regions. The selected region is called Metro East. Metro East region is mainly composed of residential consumers. Unlike commercial and industrial consumers, the residential ones are not following a working schedule. That's why it makes them to behave differently and more randomly comparing to the other two. This means more complicated demand to forecast. This is the main reason that Metro

East is selected to be studied on this chapter. One of the main components of this chapter is to introduce the methods that have been used for pre-processing of input data. The pre-processing stages include data resolution adjustment, replacement of missing data, removing outliers, clustering and signal reconstruction. A well pre-processed set of data is critical component of any forecasting strategy. The second component of chapter 4 is to generate one day ahead and seven day ahead forecasts of Metro East electricity consumption, using three different training methods. The forecasted results are comparable to other studies done on short term load forecasting. However the author questions the accuracy of classic approach of load forecasting. Classic approach is basically what have been done in the field of load forecasting for decades, which is very similar to the works done in chapter 4. In classic approach, a method gets tested on a case study with acceptable level of accuracy. Then that method gets introduces as a very accurate tool for demand forecasting purposes. This work is showing that such accurate method cannot be accurate at all when being applied to other different case studies. Future chapters study this in further details, and come up with some guidelines on how to have accurate load forecast based on the nature of the case study in hand.

Chapter 5 applies the methods of load forecasting developed in chapter 4 onto eight different case studies. By doing this, it can be seen that there is no single method of forecasting that can be accurate for all case studies out there. Temperature sensitivity and distribution of the load data of all the regions is closely studied for fifteen years of data. A load type determination criterion is presented in Table 5. By using this table, and preparing Rayleigh, Generalised Pareto, and Generalised Extreme Value distributions of the load data under study, anyone will be able to say whether their load under study is mainly commercial, residential or industrial. The outdoor temperature is the one of the main inputs of short term electricity forecasting Same chapter shows that residential loads are having a greater temperature sensitivity comparing to the other two. The results of one day and seven day ahead forecasts of the eight regions are presented at the end of chapter 5, using two methods of neural networks and decision trees. It suggest that the two methods need to be used alternatively based on the characteristics of the case study and ambient temperature for the best output.

Chapter 6explains the system based medium term load forecasting. The approach to medium term forecasting is completely different to the one developed for short term one.Two main differences between STLF and MTLF are the availability of weather data and the forecasting objectives.Because of the nature of the weather, temperature forecasts of a year ahead are completely impossible. Also in medium term load forecasting the focus of planners is mainly on peak load and energy consumption forecasts. The forecasting method presented in this chapter is achieved by superimposing annual trend, annual seasonality and forecasted residuals by neural networks and decision trees. Similar to chapter 5, the forecasting strategy is applied to eight different case studies for comparison. It is concluded that based on the case under study the accuracy of the methods change. It also provides some advices on the best practices to perform medium load forecasting, considering the characteristics of the load.

7.4 Further Applications

7.4.1 COMMERCIALISED MAPPING TOOLS

Interpreting data from tables and figures may not be convenient for a planner and he/she may require a visual representation of the forecasted load in each region. Using the mapping tool, developed in Chapter 2, all calculated values at any time of the test year can be visualised and colour coded based on value. Figure 110, Figure 111,and Figure 112present some sample outputs of the mapping tool for the test year. This is much easier to understand and can help professionals in the electricity sector make real-time decisions without needing to decipher the complicated details behind it.

Two separate Matlab functions have been developed to generate proper maps based on the forecasted values. The codes are included in Appendix D.

An example output of the code is presented onthe figures below.

Figure 110 represents the map of Perth metropolitan regions consisting of Metro North, Metro South, Metro East and CBD regions. The colour coded bar on the right side of the figure assigns colours to loads from zero to 800 MW. This will help the planner to get an idea of the load intensity at the first glance. More detailed information about the load is also presented on the map. The planner can individually look at each region and see the forecasted value at the specific time. In this case forecasted loads at 2:30 pm of 15-Mar-2015 are 592 MW, 103.83 MW, 814.9 MW and 288.8 MW respectively, for the Metro North, Metro East, Metro South and CBD regions.

Figure 111 represents the map of Perth metropolitan regions consisting of Metro North, Metro South, Metro East and CBD regions. The planner can individually look at each region and see the forecasted value at the specific time. In this case forecasted loads on 9:00 am of 12-Jul-2010 are 547.9 MW, 91.7 MW, 755 MW and 234.3 MW respectively for Metro North, Metro East, Metro South and CBD regions.

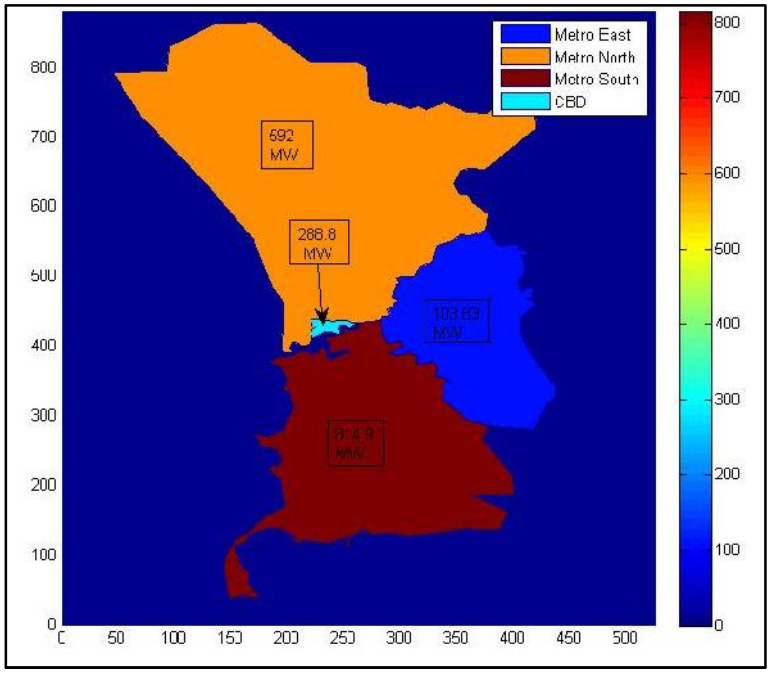


Figure 110: Spatial load forecast of Perth metropolitan region at 2:30 pm of 15-03-2010.

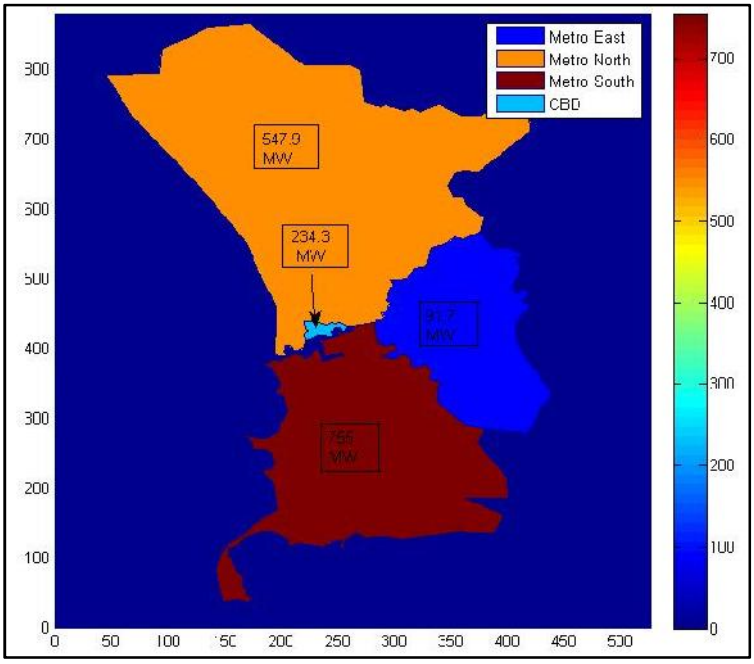


Figure 111: Spatial load forecast of Perth metropolitan region at 9:00 am of 12-07-2010.

Figure 112 represents the map of SWIS regions consisting of Country North, Country South, Country East, Country Gold fields and metropolitan region[40]. The colour coded bar on the right side of the figure assigns colour to loads from zero to 1600 mega watts. This will help the planner to get an idea of the load intensity at the first glance. More detailed information about the load is also presented on the map. The planner can individually look at each region and see the forecasted value at the specific time. In this case forecasted loads at9:00 pm of 12-Jul-2010 are 94.1 MW, 69 MW, 275 MW, 180.5 MW and 1628.9 MW respectively, for the Country North, Country East, Country South, Country Goldfields and Metropolitan regions.

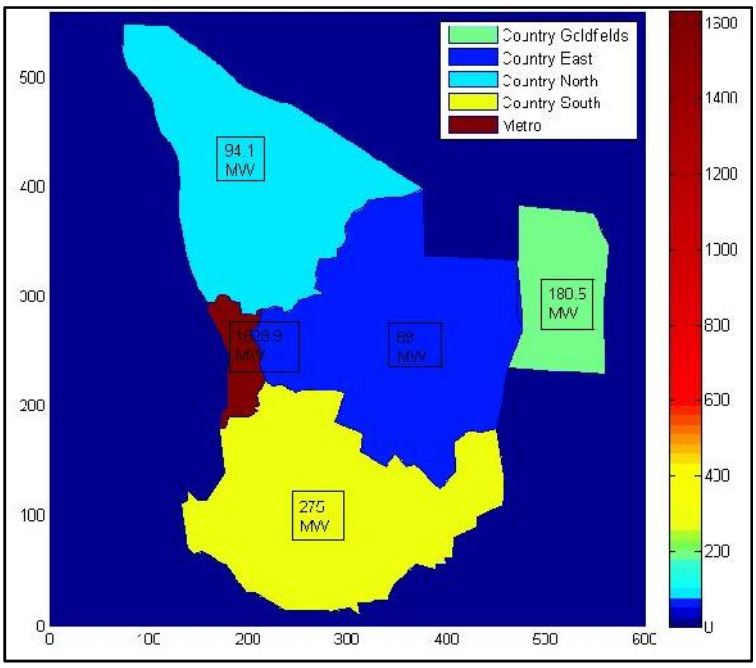


Figure 112: Spatial load forecast of SWIS region at 9:00 am of 12-07-2010.

A commercial version of such tool can be designed for Western Australia electricity network planners. Such a tool can be also designed for any other electricity network. In this way instead of seeing only one number as the total forecasted value for the network the planners can gather a detailed information of forecasted load distribution in different geographical regions.

7.4.2 DEVELOPING SPATIAL ELECTRICITY PLANNING METHODOLOGIES

7.4.2.1 Generation locations and transmission routes

Existing literature on power expansion planning usually assumes the location of generators and transmission lines as a given to simplify calculations [127]. The location of the generation plants and transmission routes can have a significant impact on power losses. To assume the location of generation plants and transmission routes in the power expansion planning is called spatial power planning[128]. Spatial load forecasting techniques developed in this work can be used as effective drive to produce reliable spatial power expansion planning strategies.

[40]The Metropolitan region consists of four regions. They are Metro North, Metro East, Metro South and CBD.

7.4.2.2 Minimising GHG emissions

Traditional methods of power planning are considering Capex and Opex reduction as their main objectives [38]. Minimising GHG emissions is now an additional objective of power planning [1], [2]. One way of achieving this is by optimising the distance of generators to the loads to reduce the transmission losses and also by harnessing the available regional sources of renewable energies and increasing their integration into the network [39]. The development of an efficient load forecasting method capable of describing the regional behaviours of the electricity consumption is a useful planning tool. Now that these efficient spatial load forecasting methods are developed, using the result of this work and combining it with traditional method of power planning, power planners can come up with multi-objective optimisation plans to reduce Capex, Opex and GHG emissions all at the same time [129].

7.4.3 BEHAVIOURAL STUDY OF LOAD PATTERNS AND THEIR EFFECTS ON FORECASTING ACCURACY

In chapter 6 the author came to a conclusion that the same forecasting methodology can have up to 10% difference in MAPE when applied to different case studies. It will be a very interesting area to be investigated in more details. A test criterion is also developed by the author that can be used to extract the dominant part of the load. It can determine whether the load is dominantly industrial, residential or commercial.

Appendix

A. BIBLIOGRAPHY

1] Q. G. Lin and G. H. Huang, "Planning of energy system management and GHG-emission control in the Municipality of Beijing—An inexact-dynamic stochastic programming model," *Energy Policy*, vol. 37, no. 11, pp. 4463–4473, Nov. 2009.

[2] G. P. Giatrakos, T. D. Tsoutsos, and N. Zografakis, "Sustainable power planning for the island of Crete," *Energy Policy*, vol. 37, no. 4, pp. 1222–1238, Apr. 2009.

[3] I. Change, "IPCC third assessment report," *Cambridge New York Cambridge Univ. Press*, pp. 365–388, 2001.

[4] M. I. BUDYKO, "The effect of solar radiation variations on the climate of the Earth," *Tellus*, vol. 21, no. 5, pp. 611–619, Oct. 1969.

[5] D. A. Lashof and D. R. Ahuja, "Relative contributions of greenhouse gas emissions to global warming," *Nature*, vol. 344, no. 6266, pp. 529–531, Apr. 1990.

[6] S. Solomon D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.) and C. Cambridge University Press United Kingdom and New York, NY, USA., "Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change." IPCC, 2007.

[7] T. K. Berntsen, I. S. A. Isaksen, G. Myhre, J. S. Fuglestvedt, F. Stordal, T. A. Larsen, R. S. Freckleton, and K. P. Shine, "Effects of anthropogenic emissions on tropospheric ozone and its radiative forcing," *J. Geophys. Res.*, vol. 102, no. D23, p. 28101, Dec. 1997.

[8] J. Hansen, M. Sato, and R. Ruedy, "Radiative forcing and climate response," *J. Geophys. Res.*, vol. 102, no. D6, p. 6831, Mar. 1997.

[9] J. T. Kiehl Trenberth, K.E., "Earth's Annual Global Mean Energy Budget ," *Am. Meteorol. Soc.*, pp. 197–209, 1997.

[10] M. Kampa and E. Castanas, "Human health effects of air pollution.," *Environ. Pollut.*, vol. 151, no. 2, pp. 362–7, Jan. 2008.

[11] J. H. Seinfeld and S. N. Pandis, *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change (Google eBook)*. John Wiley & Sons, 2012.

[12] M. R. Raupach, C. Le Quéré, G. P. Peters, and J. G. Canadell, "Anthropogenic CO2 emissions," *Nat. Clim. Chang.*, vol. 3, no. 7, pp. 603–604, Jun. 2013.

[13] P. Bousquet, P. Ciais, J. B. Miller, E. J. Dlugokencky, D. A. Hauglustaine, C. Prigent, G. R. Van der Werf, P. Peylin, E.-G. Brunke, C. Carouge, R. L. Langenfelds, J. Lathière, F. Papa, M. Ramonet, M. Schmidt, L. P. Steele, S. C. Tyler, and J. White, "Contribution of anthropogenic and natural sources to atmospheric methane variability.," *Nature*, vol. 443, no. 7110, pp. 439–43, Sep. 2006.

[14] J. A. van Aardenne, G. R. Carmichael, H. Levy, D. Streets, and L. Hordijk, "Anthropogenic

NOx emissions in Asia in the period 1990–2020," *Atmos. Environ.*, vol. 33, no. 4, pp. 633–646, Feb. 1999.

[15] J. Kim, S. Li, K.-R. Kim, A. Stohl, J. Mühle, S.-K. Kim, M.-K. Park, D.-J. Kang, G. Lee, C. M. Harth, P. K. Salameh, and R. F. Weiss, "Regional atmospheric emissions determined from measurements at Jeju Island, Korea: Halogenated compounds from China," *Geophys. Res. Lett.*, vol. 37, no. 12, p. n/a-n/a, Jun. 2010.

[16] J. W. Sutherland, "Analysis of Mechanical-Draught Counterflow Air/Water Cooling Towers," *J. Heat Transfer*, vol. 105, no. 3, p. 576, Aug. 1983.

[17] S. W. F. Dale R. Patrick, *Electrical Distribution Systems*. The Fairmont press, 1999.

[18] M. Xu, Z. Hu, J. Wu, and Y. Zhou, "A hybrid society model for simulating residential electricity consumption," *Int. J. Electr. Power Energy Syst.*, vol. 30, no. 10, pp. 569–574, Dec. 2008.

[19] B. Halvorsen and B. M. Larsen, "The flexibility of household electricity demand over time," *Resour. Energy Econ.*, vol. 23, no. 1, pp. 1–18, Jan. 2001.

[20] P. Breeze, *Power Generation Technologies*. Newnes, 2005.

[21] A. J. Pansini, *Power Transmission and Distribution*. The Fairmont Press, Inc., 2005.

[22] S. Sivanagaraju, *Electric Power Transmission and Distribution*. Pearson Education India, 2008.

[23] B.Metz, "Climate Change 2007: Energy Supply. Contribution of Working Group III to the Forth Assessment Report of the Intergovernmental Panel on Climate Change." 2007.

[24] A. Mazer, *Electric Power Planning for Regulated and Deregulated Markets*. John Wiley & Sons, 2007.

[25] M. Moghaddam and P. Bahri, "Development of a Novel Approach for Electricity Forecasting," in *IAENG Transactions on Engineering Technologies SE - 45*, vol. 247, H. K. Kim, S.-I. Ao, M. A. Amouzegar, and B. B. Rieger, Eds. Springer Netherlands, 2014, pp. 635–649.

[26] M. A. Ortega-Vazquez and D. S. Kirschen, "Estimating the Spinning Reserve Requirements in Systems With Significant Wind Power Generation Penetration," *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 114–124, Feb. 2009.

[27] *Expansion Planning for Electrical Generating Systems*, no. 241. INTERNATIONAL ATOMIC ENERGY AGENCY VIENNA, 1984.

[28] L. Parshall, D. Pillai, S. Mohan, A. Sanoh, and V. Modi, "National electricity planning in settings with low pre-existing grid coverage: Development of a spatial model and case study of Kenya," *Energy Policy*, vol. 37, no. 6, pp. 2395–2410, Jun. 2009.

[29] H. L. Willis, *Spatial Electric Load Forecasting, Second Edition*, vol. 71, no. 2. CRC, 2002.

[30] P. D. C. Wijayatunga, K. Siriwardena, W. J. L. S. Fernando, R. M. Shrestha, and R. A.

Attalage, “Strategies to overcome barriers for cleaner generation technologies in small developing power systems: Sri Lanka case study,” *Energy Convers. Manag.*, vol. 47, no. 9–10, pp. 1179–1191, Jun. 2006.

[31] *Distributed Power Generation: Planning and Evaluation*. CRC Press, 2000.

[32] J. Schlabbach and K.-H. Rofalski, *Power System Engineering: Planning, Design, and Operation of Power Systems and Equipment*. John Wiley & Sons, 2008.

[33] *Economic Market Design and Planning for Electric Power Systems*, vol. 2009. John Wiley & Sons, 2009.

[34] G. Latorre, R. D. Cruz, J. M. Areiza, and A. Villegas, “Classification of publications and models on transmission expansion planning,” *Power Syst. IEEE Trans.*, vol. 18, no. 2, pp. 938–946, 2003.

[35] M. Guk-Hyun, J. Sung-Kwan, H. Don, J. Hae-Sung, R. Heon-Su, and C. Kang-Wook, “Stochastic integrated generation and transmission planning method with gradient radar step (GRS),” 2009, pp. 1–4.

[36] E. E. El-Attar, J. Y. Goulermas, and Q. H. Wu, “Forecasting electric daily peak load based on local prediction,” *2009 IEEE Power Energy Soc. Gen. Meet.*, pp. 1–6, Jul. 2009.

[37] K. Qian, C. Zhou, M. Allan, and Y. Yuan, “Load modelling in distributed generation planning,” 2009, pp. 1–6.

[38] H. L. Willis, *Power Distribution Planning Reference Book, Second Edition*. CRC Press, 2004.

[39] K. Neuhoff, “Large-Scale Deployment of Renewables for Electricity Generation,” *Oxford Rev. Econ. Policy*, vol. 21, no. 1, pp. 88–110, Mar. 2005.

[40] H. L. Willis and J. E. D. Northcote-Green, “Spatial electric load forecasting: A tutorial review,” *Proc. IEEE*, vol. 71, no. 2, pp. 232–253, 1983.

[41] P. Strange, “History of Electric Light and Power,” *Electronics and Power*, vol. 28, no. 1. p. 112, 1982.

[42] V. K. Mehta and R. Mehta, *Principles of Power System: Including Generation, Transmission, Distribution, Switchgear and Protection*. S. Chand, 2000.

[43] K. Daley, *Electric Power Transmission*. World Technologies, 2012.

[44] J. Chow, R. J. Kopp, and P. R. Portney, “Energy resources and global development.,” *Science*, vol. 302, no. 5650, pp. 1528–31, Nov. 2003.

[45] *Renewable Energy: Sources for Fuels and Electricity*. Island Press, 1993.

[46] S. A. Kalogirou, *Solar Energy Engineering: Processes and Systems*. Academic Press, 2013.

[47] L. Freris and D. Infield, *Renewable Energy in Power Systems*. John Wiley & Sons, 2008.

[48] L. L. Grigsby, *The Electric Power Engineering Handbook*. CRC Press, 2001.

[49] G. L. Johnson, *Wind energy systems*. Prentice-Hall, 1985.

[50] J. Gordon and I. S. E. Society, *Solar Energy: The State of the Art : ISES Position Papers*. James & James, 2001.

[51] M. H. Dickson and M. Fanelli, *Geothermal energy: utilization and technology*. Earthscan, 2005.

[52] R. H. Charlier and C. W. Finkl, *Ocean Energy: Tide and Tidal Power*. Springer, 2009.

[53] D. Pimentel, *Biofuels, Solar and Wind As Renewable Energy Systems: Benefits and Risks*. Springer, 2008.

[54] S. Currie, *Hydropower*. ReferencePoint Press, 2010.

[55] V. Badescu, *Modeling Solar Radiation at the Earth's Surface: Recent Advances*. Springer, 2008.

[56] A. Goetzberger and V. U. Hoffmann, *Photovoltaic Solar Energy Generation*. Springer, 2005.

[57] *Planning and Installing Solar Thermal Systems: A Guide for Installers, Architects and Engineers*. Earthscan, 2010.

[58] A. S. Draper, *Hydropower of the Future: New Ways of Turning Water Into Energy*. Rosen Publishing Group, 2003.

[59] J. Cruz, *Ocean Wave Energy: Current Status and Future Prepectives [i.e. Perspectives]*. Springer, 2008.

[60] F. Rosillo-Calle and J. Woods, *The Biomass Assessment Handbook*. Taylor & Francis, 2012.

[61] A. C. Baker, *Tidal Power*. P. Peregrinus, 1991.

[62] Saddleback and S. E. Publishing, *Alternative Fuels*. Saddleback Educational Publishing, 2008.

[63] D. M. Tagare, *Electricity Power Generation: The Changing Dimensions*. John Wiley & Sons, 2011.

[64] A. J. Pansini and K. D. Smalling, *Guide to Electric Power Generation*. Fairmont Press, Incorporated, 2002.

[65] N. Morris, *Nuclear Power*. Smart Apple Media, 2006.

[66] T. H. Adams, *Nuclear Energy: Power from the Atom*. Crabtree Publishing Company, 2010.

[67] J. Arrillaga, *High Voltage Direct Current Transmission*. IET, 1998.

[68] T. Ackermann, G. Andersson, and L. Söder, "Distributed generation: a definition," *Electr. Power Syst. Res.*, vol. 57, no. 3, pp. 195–204, Apr. 2001.

[69] E. Lakervi and E. J. Holmes, *Electricity Distribution Network Design*. IET, 1995.

[70] A. J. Pansini, *Guide To Electrical Power Distribution Systems*. Fairmont Press, 2005.

[71] P. Holtedahl and F. L. Joutz, "Residential electricity demand in Taiwan," *Energy Econ.*, vol. 26, no. 2, pp. 201–224, Mar. 2004.

[72] J. C. Lam, "Residential sector air conditioning loads and electricity use in Hong Kong," *Energy Convers. Manag.*, vol. 41, no. 16, pp. 1757–1768, Nov. 2000.

[73] J. C. Lam, "Energy analysis of commercial buildings in subtropical climates," *Build. Environ.*, vol. 35, no. 1, pp. 19–26, Jan. 2000.

[74] E. Worrell, R. F. A. Cuelenaere, K. Blok, and W. C. Turkenburg, "Energy consumption by industrial processes in the European Union," *Energy*, vol. 19, no. 11, pp. 1113–1129, Nov. 1994.

[75] F. Elkarmi and N. Abu Shikhah, "Electricity Demand Forecasting:," *Int. J. Product. Manag. Assess. Technol.*, vol. 2, no. 1, pp. 1–19, 2014.

[76] S. M. Singh, *Electric Power Generation Transmission And Distribution 2Nd Ed*. Prentice-Hall Of India Pvt. Limited, 2010.

[77] N. Amjady, "Short-term hourly load forecasting using time-series modeling with peak load estimation capability," *IEEE Trans. Power Syst.*, vol. 16, no. 3, pp. 498–505, 2001.

[78] H. K. Alfares and M. Nazeeruddin, "Electric load forecasting: Literature survey and classification of methods," *Int. J. Syst. Sci.*, vol. 33, no. 1, pp. 23–34, Jan. 2002.

[79] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression.," *Neural Netw.*, vol. 17, no. 1, pp. 113–26, Jan. 2004.

[80] H. K. Kim, S.-I. Ao, M. A. Amouzegar, and B. B. Rieger, Eds., *IAENG Transactions on Engineering Technologies*, vol. 247. Dordrecht: Springer Netherlands, 2014.

[81] I. Moghram and S. Rahman, "Analysis and Evaluation of Five Short-Term Load Forecasting Techniques," *Power Eng. Rev. IEEE*, vol. 9, no. 11, pp. 42–43, 1989.

[82] F. J. Nogales, J. Contreras, A. J. Conejo, and R. Espinola, "Forecasting next-day electricity prices by time series models," *IEEE Trans. Power Syst.*, vol. 17, no. 2, pp. 342–348, May 2002.

[83] J. W. Taylor and P. E. McSharry, "Short-Term Load Forecasting Methods: An Evaluation Based on European Data," *Power Syst. IEEE Trans.*, vol. 22, no. 4, pp. 2213–2219, 2007.

[84] B. R. Szkuta, L. A. Sanabria, and T. S. Dillon, "Electricity price short-term forecasting using artificial neural networks," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 851–857, 1999.

[85] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: a review and evaluation," *Power Syst. IEEE Trans.*, vol. 16, no. 1, pp. 44–55, 2001.

[86] M. H. Hassoun, *Fundamentals of Artificial Neural Networks*. MIT Press, 1995.

[87] L. A. Zadeh, G. J. Klir, and B. Yuan, *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers*. World Scientific, 1996.

[88] D. K. Ranaweera, N. F. Hubele, and G. G. Karady, "Fuzzy logic for short term load forecasting," *Int. J. Electr. Power Energy Syst.*, vol. 18, no. 4, pp. 215–222, May 1996.

[89] D. K. Ranaweera, N. F. Hubele, and G. G. Karady, "Fuzzy logic for short term load forecasting," *Int. J. Electr. Power Energy Syst.*, vol. 18, no. 4, pp. 215–222, 1996.

[90] P.-F. Pai and W.-C. Hong, "Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms," *Electr. Power Syst. Res.*, vol. 74, no. 3, pp. 417–425, Jun. 2005.

[91] M.-G. Zhang, "Short term load forecasting based on support vector machines regression ." Guangzhou, China, p. 4310, 2005.

[92] J. R. Stewart, "Integration of wind capacity into an isolated thermal power system," in *6th International Conference on Advances in Power System Control, Operation and Management. Proceedings. APSCOM 2003*, 2003, vol. 2003, pp. 722–725.

[93] Western Power, "Transmission and distribution 2010 annual planning report," 2010.

[94] H. L. Willis, *Spatial Electric Load Forecasting, second edition revised and expanded*. Marcel Dekker, 2002.

[95] S. F. Methods, "Comparison and Selection of Spatial Forecast Methods."

[96] E. M. Carreno and A. Padilha-Feltrin, "Spatial Electric Load Forecasting Using a Local Movement Approach," 2009, pp. 1–6.

[97] L. Liu and Z. Ma, "A combined practical short term spatial load forecasting method in distribution system," in *18th International Conference and Exhibition on Electricity Distribution (CIRED 2005)*, 2005, no. June, pp. v5-91-v5-91.

[98] H. Edelsbrunner, *Geometry and Topology for Mesh Generation*. Cambridge University Press, 2001.

[99] R. Weron, *Modeling and forecasting electricity loads and prices: statistical approach*. Wiley, 2006.

[100] H. K. Temraz, M. M. a. Salama, and A. Y. Chikhani, "Review of electric load forecasting methods," in *IEEE Canadian Conference on Electrical and Computer Engineering*, 1997, vol. 1, pp. 289–292.

[101] H. Hahn, S. Meyer-Nieberg, and S. Pickl, "Electric load forecasting methods: Tools for decision making," *Eur. J. Oper. Res.*, vol. 199, no. 3, pp. 902–907, Dec. 2009.

[102] A. Sharaf, T. Lie, and H. Gooi, "A neural network based short term load forecasting model," in *IEEE Canadian Conference on Electrical and Computer Engineering*, 1993, pp. 325–328.

[103] N. Amral, D. King, and C. S. Ozveren, "Application of artificial neural network for short term load forecasting," *43rd Int. Univ. Power Eng. Conf.*, pp. 1–5, Sep. 2008.

[104] S. Munkhjargal and V. Z. Manusov, "Artificial neural network based short-term load forecasting," in *The 8th Russian-Korean International Symposium on Science and Technology*, 2004, vol. 1, pp. 262–264.

[105] T. Peng, N. Hubele, and G. Karady, "An adaptive neural network approach to one-week ahead load forecasting," *IEEE Trans. Power Syst.*, vol. 8, no. 3, pp. 1195–1203, 1993.

[106] E. Lobato, A. Ugedo, and R. Rouco, "Decision Trees Applied to Spanish Power Systems Applications," *IEEE 9th Int. Conf. Probabilistic Methods Appl. to Power Syst.*, pp. 1–6, Jun. 2006.

[107] C. Tranchita and A. Torres, "Soft computing techniques for short term load forecasting," in *Power Systems Conference and Exposition, 2004. IEEE PES*, 2004, pp. 497–502.

[108] H. Chen, Y. Du, and J. N. Jiang, "Weather sensitive short-term load forecasting using knowledge-based ARX models," in *IEEE Power Engineering Society General Meeting*, 2005, pp. 190–196.

[109] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.

[110] J. Franklin, "The elements of statistical learning: data mining, inference and prediction," *Math. Intell.*, vol. 27, no. 2, pp. 83–85, Nov. 2008.

[111] T. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization," *Mach. Learn.*, vol. 22, pp. 1–22, 2000.

[112] R. Weron, *Modeling and forecasting electricity loads and prices A Statistical Approach*. 2006.

[113] "Australian bureau of meteorology." [Online]. Available: www.bom.gov.au.

[114] A. S. AlFuhaid, M. A. El-Sayed, and M. S. Mahmoud, "Cascaded artificial neural networks for short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 12, no. 4, pp. 1524–1529, 1997.

[115] T. W. S. Chow and C. T. Leung, "Neural network based short-term load forecasting using weather compensation," *IEEE Trans. Power Syst.*, vol. 11, no. 4, pp. 1736–1742, 1996.

[116] K. Y. Lee, Y. T. Cha, and J. H. Park, "Short-term load forecasting using an artificial neural network," *IEEE Trans. Power Syst.*, vol. 7, no. 1, pp. 124–132, 1992.

[117] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: a review and evaluation," *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 44–55, 2001.

[118] D. Bunn and E. D. Farmer, "Comparative models for electrical load forecasting," Jan. 1985.

[119] "IEEE Xplore Full-Text HTML : Weather sensitive short-term load forecasting using knowledge-based ARX models." [Online]. Available: http://ieeexplore.ieee.org/xpls/icp.jsp?arnumber=1489086. [Accessed: 19-Apr-2015].

[120] A. Spanos, *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data*. Cambridge University Press, 1999.

[121] A. Lyon, "Why are Normal Distributions Normal?," *Br. J. Philos. Sci.*, vol. 65, no. 3, pp. 621–

649, Sep. 2013.

[122] S. Africa, "Statistics of Extremes and Applications : An Introduction," p. 490, 2014.

[123] D. K. Ranaweera, G. G. Karady, and Richard. G. Farmer, "Economic impact analysis of load forecasting," *IEEE Trans. Power Syst.*, vol. 12, no. 3, pp. 1388–1392, 1997.

[124] H. Hahn, S. Meyer-Nieberg, and S. Pickl, "Electric load forecasting methods: Tools for decision making," *Eur. J. Oper. Res.*, vol. 199, no. 3, pp. 902–907, Dec. 2009.

[125] S. Mirasgedis, Y. Sarafidis, E. Georgopoulou, D. Lailas, M. MoschovitsS, F. Karagiannis, and D. Papakonstantinou, "Models for mid-term electricity demand forecasting incorporating weather influences," *Energy*, vol. 31, no. 2–3, pp. 208–227, Feb. 2006.

[126] J. V. Ringwood, D. Bofelli, and F. T. Murray, "Forecasting Electricity Demand on Short, Medium and Long Time Scales Using Neural Networks," *J. Intell. Robot. Syst.*, vol. 31, no. 1–3, pp. 129–147.

[127] J. Shu, L. Wu, L. Zhang, and B. Han, "Spatial Power Network Expansion Planning Considering Generation Expansion," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 1815–1824, Jul. 2015.

[128] J. Shu, L. Wu, Z. Li, M. Shahidehpour, L. Zhang, and B. Han, "A New Method for Spatial Power Network Planning in Complicated Environments," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 381–389, Feb. 2012.

[129] C. Unsihuay-Vila, J. W. Marangon-Lima, A. C. Zambroni de Souza, and I. J. Perez-Arriaga, "Multistage expansion planning of generation and interconnections with sustainable energy development criteria: A multiobjective model," *Int. J. Electr. Power Energy Syst.*, vol. 33, no. 2, pp. 258–270, Feb. 2011.

[130] "timeanddate.com." [Online]. Available: http://www.timeanddate.com/.

[131] ABARE, "Australian energy resource assessment," 2010.

[132] "Australian bureau of statistics." [Online]. Available: http://www.abs.gov.au/.

[133] J. E. Penner, *Aviation and the Global Atmosphere: A Special Report of IPCC Working Groups I and III in Collaboration with the Scientific Assessment Panel to the Montreal Protocol on Substances that Deplete the Ozone Layer*. Cambridge University Press, 1999.

[134] R. Khandekar, "The prevalence and causes of blindness in the Sultanate of Oman: the Oman Eye Study (OES)," *Br. J. Ophthalmol.*, vol. 86, no. 9, pp. 957–962, Sep. 2002.

[135] G. Chicco, R. Napoli, and F. Piglione, "Load pattern clustering for short-term load forecasting of anomalous days," in *IEEE Porto Power Tech Proceedings*, 2001, vol. vol.2.

[136] M. Meng, J. Lu, and W. Sun, "Short-Term Load Forecasting Based on Ant Colony Clustering and Improved BP Neural Networks," in *International Conference on Machine Learning and Cybernetics*, 2006, pp. 3012–3015.

[137] A. Inc, *ASHRAE Standard 55—thermal environmental conditions for human occupancy*.

ASHRAE Inc, 2013.

Appendix

B. Public Holidays in Western Australia

Western Australia has 13 days of public holidays during each year. The following is a summarised list of those holidays and a short description of each [130][41]:

New Year's Day: The first day of January is the beginning of the year, according to Gregorian calendar.

Australia Day: This day is celebrated each year on 26th of January in Australia.

Labor Day: The first Monday of March is celebrated as Labor Day in Western Australia. The date is different in the other states.

Good Friday: As a day before Easter Sunday, this day is a public holiday across Australia. Christians commemorate the crucifixion of Jesus on this day.

Easter Saturday: As the day before Easter Sunday and the day after Good Friday, this day is a public holiday across Australia.

Easter Sunday: Christians celebrate this day to commemorate the resurrection of Jesus after his crucifixion on Good Friday.

Easter Monday: As the day after Easter Sunday, this day is a public holiday in Australia.

Anzac Day: The 25th of April each year is considered the anniversary of Australian troops landing on the Gallipoli Peninsula, Turkey during the World War I.

Foundation Day: The first Monday of June is celebrated as foundation day in Western Australia to commemorate the foundation of the Swan River Colony.

Queen's birthday: As a constitutional monarchy, Australia celebrates the Queen's birthday each year. The date is usually different in the other states. In Western Australia, this day is celebrated on the last Monday of September or on the first Monday of October.

Christmas Day: The 25th of December is the day that Christians celebrate the birthday of Jesus.

Boxing Day: The 26th of December is celebrated as the day after Christmas Day.

New Year's Eve: The 31st of December is one day before New Year's Day, according to the Gregorian calendar.

The following tables show the distribution of public holidays in Western Australia from 1995 to 2011.

Table 16:Western Australian public holidays from 1995 to 1999 inclusive.

[41]All information about Western Australian holidays has been extracted from this source.

	1995	1996	1997	1998	1999
New Year's Day	1-Jan	1-Jan	1-Jan	1-Jan	1-Jan
Australia Day	26-Jan	26-Jan	26-Jan	26-Jan	26-Jan
Labor Day	6-Mar	4-Mar	4-Mar	2-Mar	1-Mar
Good Friday	14-Apr	5-Apr	28-Ma	10-Apr	2-Apr
Easter Saturday	15-Apr	6-Apr	29-Mar	11-Apr	3-Apr
Easter Sunday	16-Apr	7-Apr	30-Mar	12-Apr	4-Apr
Easter Monday	17-Apr	8-Apr	31-Mar	13-Apr	5-Apr
Anzac Day	25-Apr	25-Apr	25-Apr	25-Apr	25-Apr
Foundation Day	5-Jun	3-Jun	2-Jun	1-Jun	7-Jun
Queen's Birthday	2-Oct	30-Sep	21-Sep	28-Sep	27-Sep
Christmas Day	25-Dec	25-Dec	25-Dec	25-Dec	25-Dec
Boxing Day	26-Dec	26-Dec	26-Dec	26-Dec	26-Dec
New Year's Eve	31-Dec	31-Dec	31-Dec	31-Dec	31-Dec

Table 17:Western Australian public holidays from 2000 to 2004 inclusive.

	2000	2001	2002	2003	2004
New Year's Day	1-Jan	1-Jan	1-Jan	1-Jan	1-Jan
Australia Day	26-Jan	26-Jan	26-Jan	26-Jan	26-Jan
Labor Day	6-Mar	5-Mar	4-Mar	3-Mar	1-Mar
Good Friday	21-Apr	13-Apr	29-Mar	18-Apr	9-Apr
Easter Saturday	22-Apr	14-Apr	30-Mar	19-Apr	10-Apr
Easter Sunday	23-Apr	15-Apr	31-Mar	20-Apr	11-Apr
Easter Monday	24-Apr	16-Apr	1-Apr	21-Apr	12-Apr
Anzac Day	25-Apr	25-Apr	25-Apr	25-Apr	25-Apr
Foundation Day	5-Jun	4-Jun	3-Jun	2-Jun	7-Jun
Queen's Birthday	2-Oct	1-Oct	30-Sep	29-Sep	27-Sep
Christmas Day	25-Dec	25-Dec	25-Dec	25-Dec	25-Dec
Boxing Day	26-Dec	26-Dec	26-Dec	26-Dec	26-Dec
New Year's Eve	31-Dec	31-Dec	31-Dec	31-Dec	31-Dec

Table 18:Western Australian public holidays from 2005 to 2009 inclusive.

	2005	2006	2007	2008	2009
New Year's Day	1-Jan	1-Jan	1-Jan	1-Jan	1-Jan
Australia Day	26-Jan	26-Jan	26-Jan	26-Jan	26-Jan
Labor Day	7-Mar	6-Mar	5-Mar	3-Mar	2-Mar
Good Friday	25-Mar	14-Apr	6-Apr	21-Mar	10-Apr
Easter Saturday	26-Mar	15-Apr	7-Apr	22-Mar	11-Apr
Easter Sunday	27-Mar	16-Apr	8-Apr	23-Mar	12-Apr
Easter Monday	28-Mar	17-Apr	9-Apr	24-Mar	13-Apr
Anzac Day	25-Apr	25-Apr	25-Apr	25-Apr	25-Apr

Foundation Day	6-Jun	5-Jun	4-Jun	2-Jun	1-Jun
Queen's Birthday	26-Sep	2-Oct	1-Oct	29-Sep	28-Sep
Christmas Day	25-Dec	25-Dec	25-Dec	25-Dec	25-Dec
Boxing Day	26-Dec	26-Dec	26-Dec	26-Dec	26-Dec
New Year's Eve	31-Dec	31-Dec	31-Dec	31-Dec	31-Dec

Table 19:Western Australian public holidays from 2010 to 2011 inclusive.

	2010	**2011**
New Year's Day	1-Jan	1-Jan
Australia Day	26-Jan	26-Jan
Labor Day	1-Mar	7-Mar
Good Friday	2-Apr	22-Apr
Easter Saturday	3-Apr	23-Apr
Easter Sunday	4-Apr	24-Apr
Easter Monday	5-Apr	25-Apr
Anzac Day	25-Apr	25-Apr[42]
Foundation Day	7-Jun	6-Jun
Queen's Birthday	27-Sep	28-Oct
Christmas Day	25-Dec	25-Dec
Boxing Day	26-Dec	26-Dec
New Year's Eve	31-Dec	31-Dec

[42]In 2011, Anzac Day and Easter Monday were observed on the same day, on April 25.

Appendix

C. AUSTRALIA'S DISTRIBUTION FIGURES

The following are the distribution figures of power stations, resources, weather, and population across Australia, extracted from [113], [131], [132].

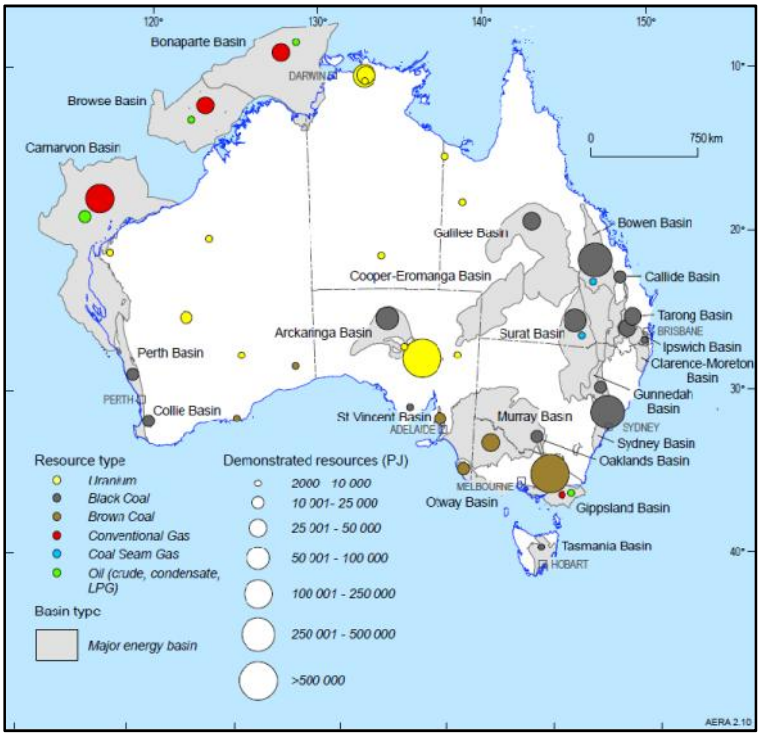


Figure 113:Major non-renewable energy resources distribution in 2010(extracted from [131]).

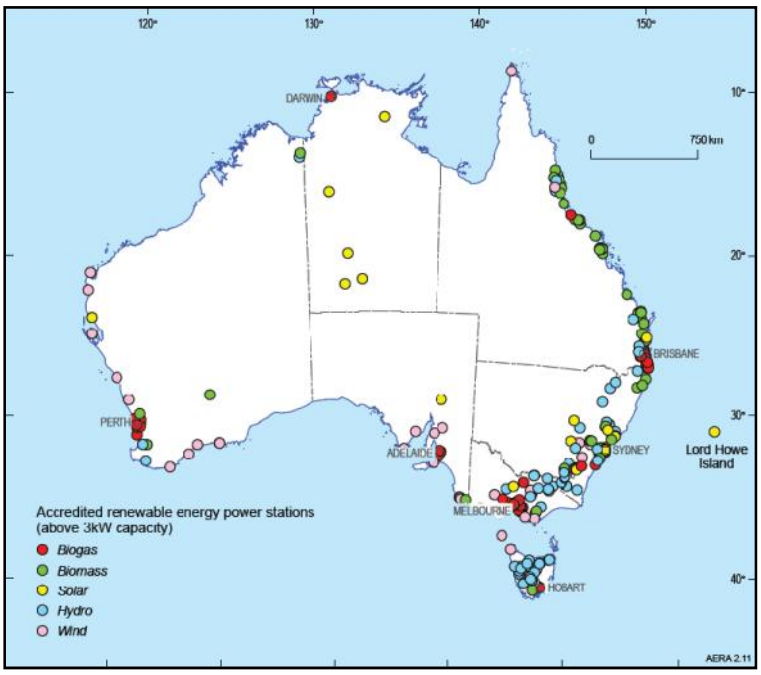


Figure 114: Renewable energy power stations distribution (more than 3kW capacity) in 2010(extracted from [131]).

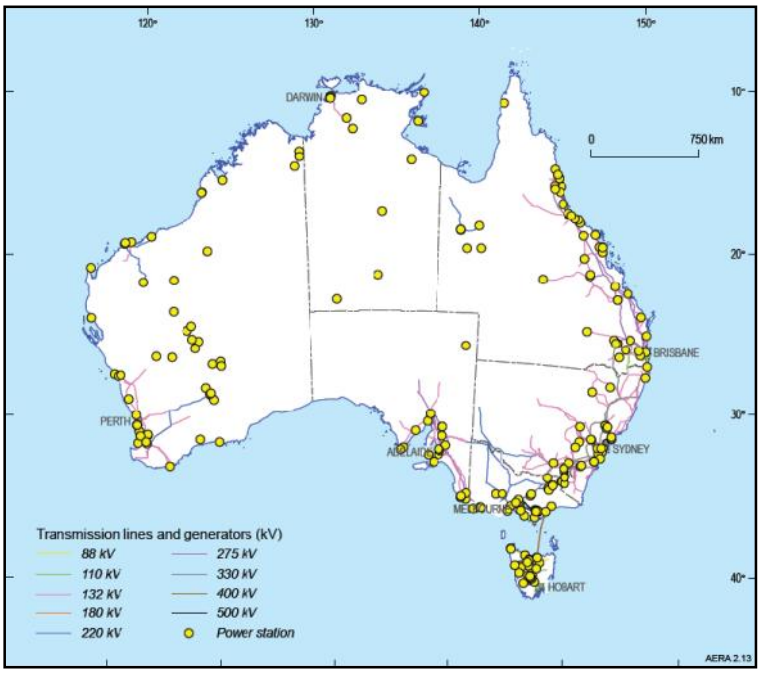


Figure 115: Australia's electricity infrastructure in 2010 (extracted from [131]).

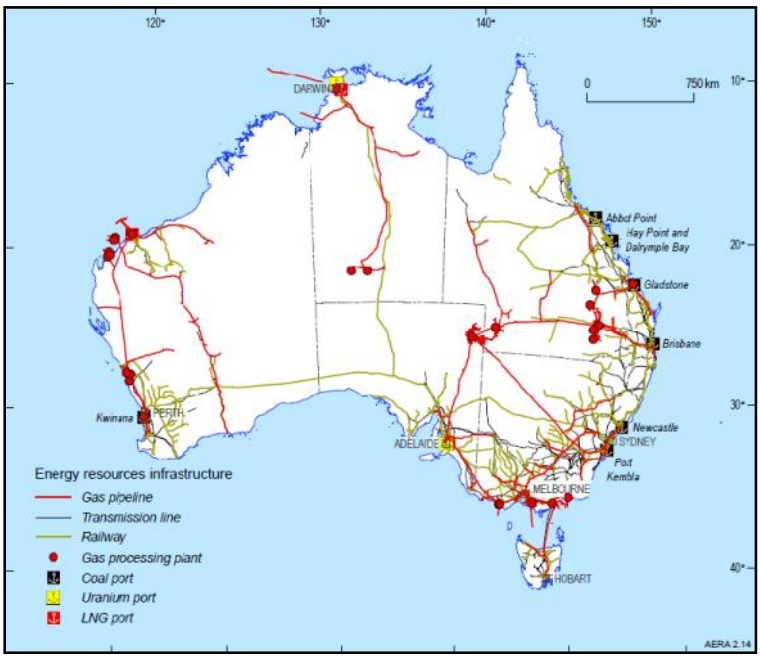


Figure 116:Australia's non-renewable energy resource infrastructure in 2010 (extracted from [131]).

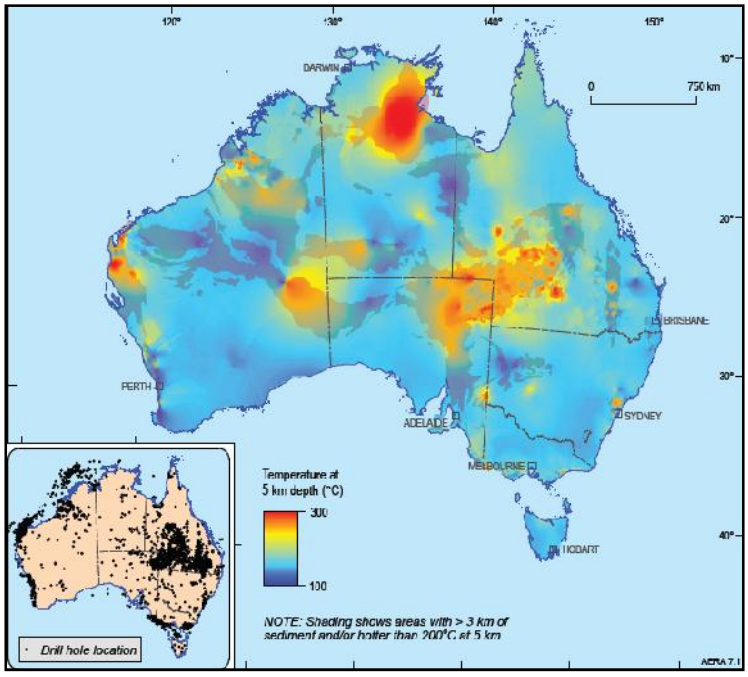


Figure 117:Australia's geothermal energy distribution (based on data from more than 5000 petroleum and water boreholes) in 2010(extracted from [131]).

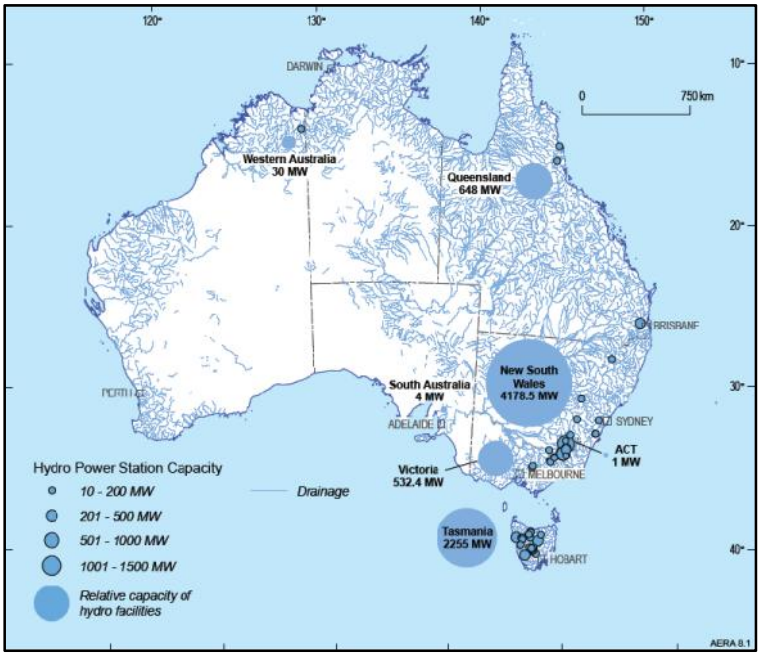


Figure 118: Australia's major operating hydro power stations in 2010 (extracted from [131]).

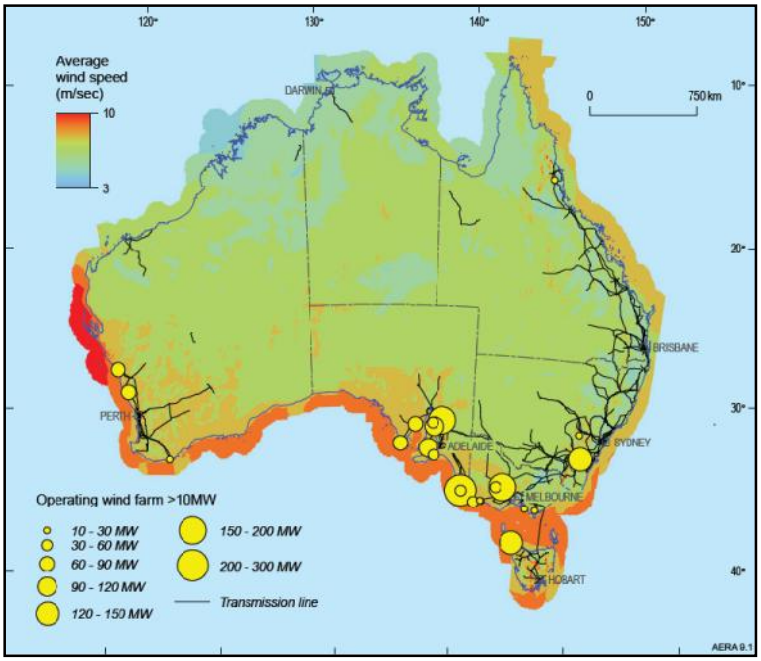


Figure 119:Distribution of Australia's wind resources and major farms (more than 10 MW capacity) in 2010(extracted from [131]).

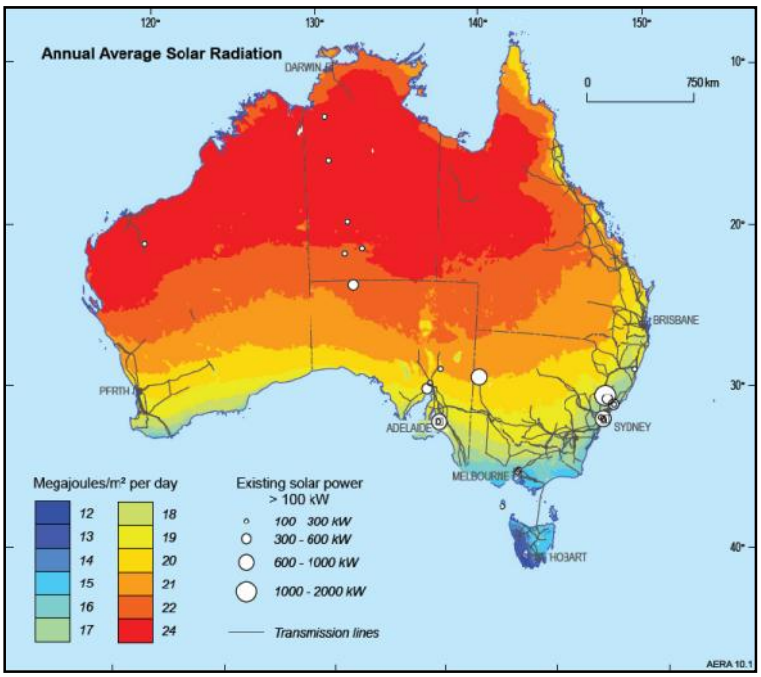


Figure 120**:** Annual average solar radiation and generation units of more than 10 kW capacity in 2010 (extracted from [131]).

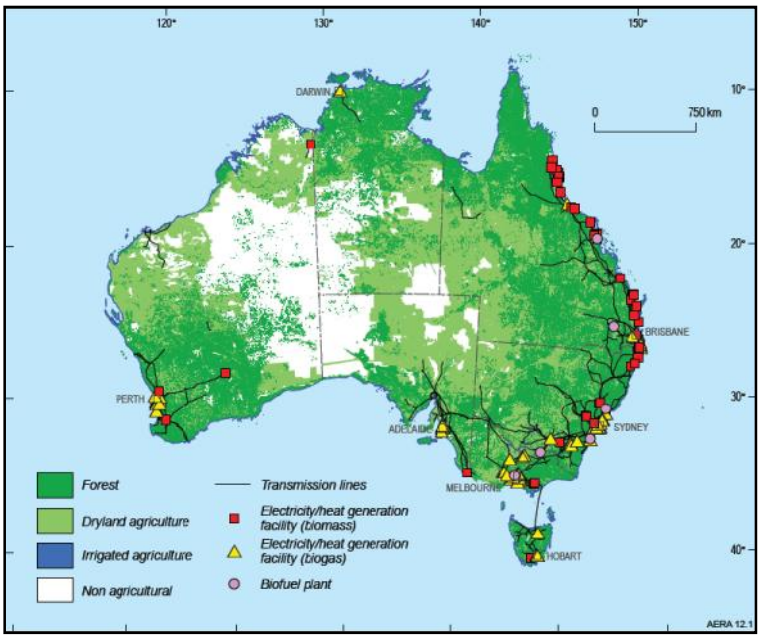


Figure 121**:** Land use and bioenergy facilities in 2010 (extracted from [131]).

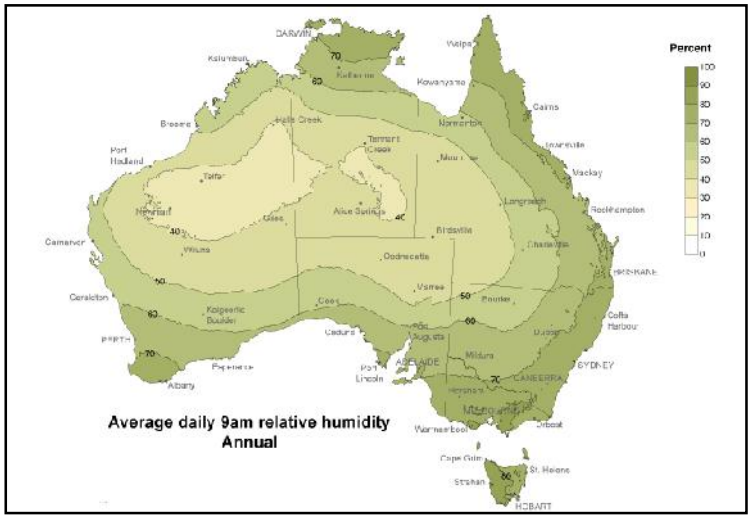


Figure 122**:** Average annual relative humidity at 9 am (based on data from 1976 to 2005) (extracted from [113]).

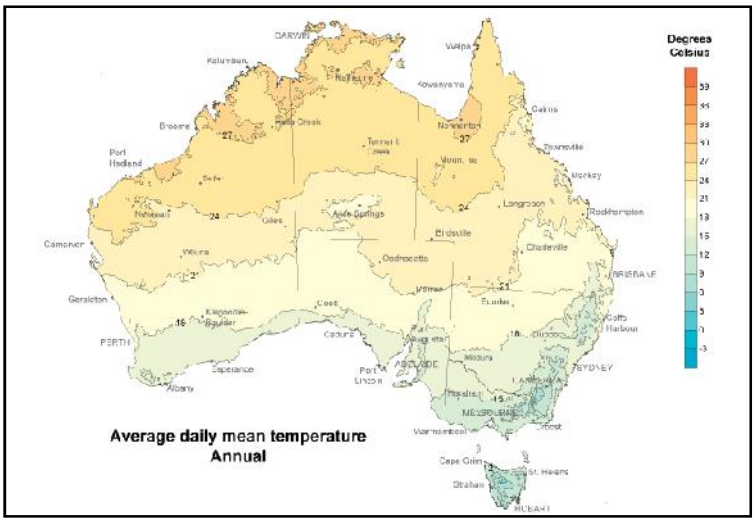


Figure 123**:** Average annual temperature (based on data from 1961 to 1990) (extracted from [113]).

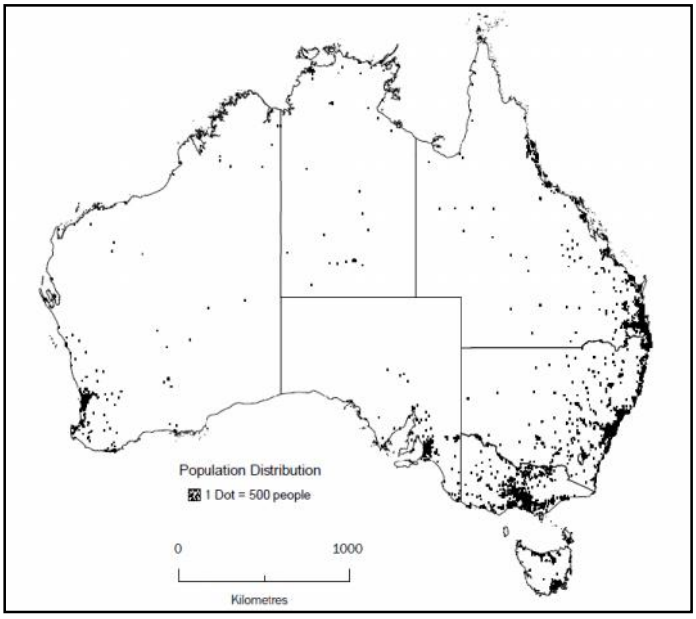


Figure 124:Population distribution 2001 (extracted from [132]).

Appendix

D. SAMPLE CODES

This Appendix contains some of the sample codes that have been developed for this research. The codes have been developed in Matlab unless otherwise stated.[43]

i. DATA RESOLUTION ADJUSTMENT

The resolution of weather data of seven different stations has been adjusted to half hourly using the following code.

```
%-------------------------------------------------------------------
-------
%Title: changing the resolution of temperature and humidity data
%-------------------------------------------------------------------
-------
%Status: has been tested and is working properly-modified on 22-06-
2011 and
%-------------------------------------------------------------------
-------
%Date: 27-04-2011
%-------------------------------------------------------------------
-------
%Usage: The resolution of BOM data are hourly and we are after half
an
%hourly resolution. Because of the nature of temperatue and humidity
fist
%order hold is applied in each half an hour period.
%-------------------------------------------------------------------
-------
%How to use this code: Matrix A is a cell input from Bom database
which is
%usually a n*2 matrix with hourly resolution of data. Matrix B as an
output
%of this code is a 2n*2 matrix with half an hourly resolution data
of
%temperature and humidity. First order hold has been used for
estimation of
%half an hourly data. Previous value is placed in the places in
which
%temperature and,humidity were not available.
%-------------------------------------------------------------------
-------
clc;
clear A b B Number_of_Rows Number_of_Columns Row_Counter
Column_Counter i j;
```

[43]Matlab is the main software that has been used in this study. A couple of codes have also been developed in VBA and SAS environment.

```
A=cell(1,1);
disp('Please copy and paste Matrix A and type return when
finished.');
disp('Hint: You may use Ctrl+Shift+V to paste data from excel data
sheet');
keyboard;
[Number_of_Rows , Number_of_Columns] = size(A);
%Date_Array_Double = datevec(A(:,1));
b= cell(0, Number_of_Columns);
%this part will make a matrix twice the size of the input matrix
with real
%values on odd rows and repeated values on even rows
for Row_Counter=1:Number_of_Rows
  b=[b; A(Row_Counter,:)];
  b=[b; A(Row_Counter,:)];
end
[i,j]=size(b);
B=zeros(i,Number_of_Columns);
%this part just replace the non number values with nan and replace
the
%remaining with its equivalent number......cell----> num
for Row_Counter=1:i
  for Column_Counter=1:Number_of_Columns
    if length(double(cell2mat(b(Row_Counter,Column_Counter))))==1;
      B(Row_Counter,Column_Counter)=
cell2num(b(Row_Counter,Column_Counter));
    else
      B(Row_Counter,Column_Counter)= nan;
    end
  end
end
%this part replace even rows with an average value using foh
for Row_Counter=1:Number_of_Rows-1
  for Column_Counter=1:Number_of_Columns
    if B(Row_Counter*2-1)== nan
      B(Row_Counter*2,Column_Counter)=
B(Row_Counter*2+1,Column_Counter)
    elseif B(Row_Counter*2+1)== nan
        B(Row_Counter*2,Column_Counter)= B(Row_Counter*2-
1,Column_Counter)
      else
        B(Row_Counter*2,Column_Counter)=mean([B(Row_Counter*2-
1,Column_Counter),B(Row_Counter*2+1,Column_Counter)]);
    end
  end
end
```

ii. Error Messages and Missing Data Replacement

```
%-------------------------------------------------------------
-------------
%Title: no data to nan
%-------------------------------------------------------------
-------------
%Status: has been tested and is working properly
%-------------------------------------------------------------
-------------
%Date: 10-11-2011
%-------------------------------------------------------------
-------------
%Note: This code will get an input matrix of type cell and it
will find all
%the missing values and error messages with nan. The output
matrix is also
%cell type
%-------------------------------------------------------------
-------------
clc
clear A
A = cell(1,1);
disp('Please copy and paste Matrix A and type return when
finished.');
disp('Hint: You may use Ctrl+Shift+V to paste data from excel
data sheet');
keyboard;
[Number_of_Rows , Number_of_Columns] = size(A);
for i=1:Number_of_Rows
  for j=2:Number_of_Columns %1st column contains date and time
data
    if length(double(cell2mat(A(i,j))))~=1
      A(i,j)=num2cell(nan);
    end
  end
end
```

iii. Converting Date Values to Numbers

```
%-------------------------------------------------------------
-------------
%Title: date to vector
%-------------------------------------------------------------
-------------
%Status: has been tested and is fully functional
%-------------------------------------------------------------
-------------
%Date: 11-11-2011
```

```
%Modified: 12-11-2011
%-------------------------------------------------------------
-------------
%Note: This code will get an input matrix of type cell and it
will convert
%all the date and time variables into numbers and will add new
columns for
%them. The output will be a double array(it will also
recognise the
%weekday)
%Example: this code will convert 11-Jan-2003 17:30:00 to
%[2003,01,11,17.5,7]...the last element is the weekday 7 means
saturday
%-------------------------------------------------------------
-------------
% X1 year
% X2 Month
% X3 Day
% X4 Hour
% X5 Weekday
disp('The input matrix should not contain any cell with error
message or value')
disp('Hint: You can remove them by using nodata2nan code')
a=input('Please specify the input matrix:');
[X1 X2 X3 X4 temp2 temp3]=datevec(a(:,1));
X4=X4+temp2/60;
for counter=1:length(a)
  X5(counter,1)=weekday([num2str(X1(counter)),'-
',num2str(X2(counter)),'-',num2str(X3(counter))]);
end
%making the double type output matrix, with date vector
instead of dates
%and also the weekdays in the 4th column
b=[X1, X2, X3, X4,X5, cell2num(a(:,2:end))];
```

iv. Missing Data Point Estimation

```
%-------------------------------------------------------------
-------------
%Title: short-term estimation for missing transformers load
data
%-------------------------------------------------------------
-------------
%Status: has been tested and is working properly
%-------------------------------------------------------------
-------------
%Date: 28-02-2011
```

```
%Modified: 11-11-2011
%-------------------------------------------------------------
-------------
%Usage: Among SWIS transformers load data there are some
places in which no
%good data is available for calculation. The main reason for
this is
%usually data transmission failure, which means that a
transmiter was not
%able to send data to the control room in that period of time.
The
%mentioned period of time varies from several minutes to few
days. Another
%source of this can be from outliers removal step.This
%code will compensate these faults by a reasonable
shortforecasting method.
%-------------------------------------------------------------
-------------
%Method description: According to the available literature on
electricity
%load consumption behaviors[Asber D, Lefebvre S, Asber J,
Nonparametric...]
% [Seppala A. load research and load estimation],
%[Hyndman RJ, Fan S. Density forecasting for long-term peak
demand...]
%and also the pactical methods which are using in the
electricty industries
%a transformer load in a specific time of a day is closely
related to
%previous and also future weeks load data at the same time of
the same day
%of the week. Using the existed autocorrelation in load data
this method
%is going to use as much as available data in previous and
future weeks 0f
%load data at the same time of the day to replace the missed
data with
%fairly accurate estimated ones.
%-------------------------------------------------------------
-------------
% Method accuracy level: Some important variables like ambient
temperature
% and humidity level have not been considered directly in this
short time
% estimating method, and this method is only based on
historical and future
% load data. The reason is that couple of hours of
transformers load data is
```

```
% really a small value in comparison with the whole region
electricity load
% consumption and using load autocorrelation in this case is
reasonable.
%-----------------------------------------------------------
-------------
%How to use this code?: Matrix A is the input matrix which may
include some
% NAN DATA in it. The input is matrix A with date and time
information in
%the first four column, temperature in 5th column, relative
humidity in
%6th column and 7th column contains load information. The
% output of this code will be matrix B which has the same size
as matrix A
% but all of its cells are real numbers. Nans will replaced
with their
%proper estimation in the new array.
%THIS IS VERY IMPORTANT TO KNOW that this m file is to help
out the user to
%fill the missed cells as quick as possible. This is not a
fully automatic
%code and you will need to copy and paste it section by
section in Matlab
%workspace and feed it manually.
%-----------------------------------------------------------
-------------
% spline will be used for both interapolation and
extrapolation purposes
clc;
clear A B Number_of_Rows Number_of_Columns Row_Counter
Column_Counter x y i check
Number_of_Rows = input('How many Rows does your input matrix
have? :');
Number_of_Columns = input('How many columns does your input
matrix have? :');
A = zeros(Number_of_Rows, Number_of_Columns);
disp('Please copy and paste Matrix A and type return when
finished.');
disp('Array size will automatically change if it''s bigger
than the entered values.')
disp('Hint: You may use Ctrl+Shift+V to paste data from excel
data sheet');
keyboard;
B = A;
[Number_of_Rows , Number_of_Columns] = size(B);
missing_data=true;
test=0;
```

```
while missing_data
for Column_Counter=8:8
  if test==1
    missing_data=false;
  end
  test=1;
  for Row_Counter=1:Number_of_Rows
      %this part will replace missing data by spline fitting
      %maximum of four neighboring numbers will be used for
fitting
      %336 is the number of intervals between two neighbors
"this
      %number comes from half an hourly distribution of data
over a
      %week"
      if isnan(B(Row_Counter,Column_Counter))
      clear x y
      x=[];
      x1=[];
      x2=[];
      y=[]; %load
      y1=[];%temperature
      y2=[];%relative humidity
      check=0; %number of data points to fit
      pdw=0; % weight for previous 30 minutes data point
      for i = -2:2
        if(Row_Counter+i*336>0) &&
(Row_Counter+i*336<Number_of_Rows)
          if ~isnan(B(Row_Counter+i*336,Column_Counter))
            x = [x i];
            y = [y B(Row_Counter+i*336,Column_Counter)];
            if ~isnan(B(Row_Counter+i*336,Column_Counter-2))
              y1 = [y1 B(Row_Counter+i*336,Column_Counter-2)];
              x1 = [x1 i];
            end
            if ~isnan(B(Row_Counter+i*336,Column_Counter-1))
              y2 = [y2 B(Row_Counter+i*336,Column_Counter-1)];
              x2 = [x2 i];
            end
            check = check + 1;
          end
        end
      end

      if (check > 1) %&& (Row_Counter > 9)
        B(Row_Counter,Column_Counter)= ((spline(x,y,0)));
        if length(y2)>1
```

```
          B(Row_Counter,Column_Counter-1)=
((spline(x2,y2,0)));
        elseif length(y2)==1
          B(Row_Counter,Column_Counter-1)= y2;
        end
        if length(y1)>1
          B(Row_Counter,Column_Counter-2)=((spline(x1,y1,0)));
        elseif length(y2)==1
          B(Row_Counter,Column_Counter-2)= y1;
        end
      end
      %if (check > 1) %&& (Row_Counter == 9)
       %  B(Row_Counter,Column_Counter)= (spline(x,y,0));
      %end
      if check==1
          B(Row_Counter,Column_Counter)= (y);
        if length(y2)>1
          B(Row_Counter,Column_Counter-1)=
((spline(x2,y2,0)));
        elseif length(y2)==1
          B(Row_Counter,Column_Counter-1)= y2;
        end
        if length(y1)>1
          B(Row_Counter,Column_Counter-2)=((spline(x1,y1,0)));
        elseif length(y2)==1
          B(Row_Counter,Column_Counter-2)= y1;
        end
      end
      if check==0
        disp('Cannot find enough neighbors to fit dada.');
        Row_Counter
        test=0;

      end
      end
  end
end
end
```

v. Q-Q PLOTS OF ALL THE REGIONS

```
%-------------------------------------------------------------
-------------
%Title: all regions Q-Q plots of different distributions
%-------------------------------------------------------------
-------------
%Status: has been tested and is working properly
```

```
%-------------------------------------------------------------
-------------
%Date: 02-02-2012
%-------------------------
close all
%-NP-----------------------------------------------------
figure
subplot(4,2,1)
gqqplot(NP_ready(100000:117520,8),'norm')
grid
ylabel('NP quantiles')
xlabel('Normal quantiles')
title('(a)')
subplot(4,2,2)
gqqplot(NP_ready(100000:117520,8),'gam')
grid
ylabel('NP quantiles')
xlabel('Gamma quantiles')
title('(b)')
subplot(4,2,3)
gqqplot(NP_ready(100000:117520,8),'logn')
grid
ylabel('NP quantiles')
xlabel('Lognormal quantiles')
title('(c)')
subplot(4,2,4)
gqqplot(NP_ready(100000:117520,8),'wbl')
grid
ylabel('NP quantiles')
xlabel('Weibull quantiles')
title('(d)')
subplot(4,2,5)
gqqplot(NP_ready(100000:117520,8),'gev')
grid
ylabel('NP quantiles')
xlabel('Generalized extreme value quantiles')
title('(e)')
subplot(4,2,6)
gqqplot(NP_ready(100000:117520,8),'gp')
grid
ylabel('NP quantiles')
xlabel('Generalized pareto quantiles')
title('(f)')
subplot(4,2,7)
gqqplot(NP_ready(100000:117520,8),'poiss')
grid
ylabel('NP quantiles')
xlabel('Poisson quantiles')
```

```
title('(g)')
subplot(4,2,8)
gqqplot(NP_ready(100000:117520,8),'rayl')
grid
ylabel('NP quantiles')
xlabel('Rayleigh quantiles')
title('(h)')
%-EP-------------------------------------------------
figure
subplot(4,2,1)
gqqplot(EP_ready(100000:117520,8),'norm')
grid
ylabel('EP quantiles')
xlabel('Normal quantiles')
title('(a)')
subplot(4,2,2)
gqqplot(EP_ready(100000:117520,8),'gam')
grid
ylabel('EP quantiles')
xlabel('Gamma quantiles')
title('(b)')
subplot(4,2,3)
gqqplot(EP_ready(100000:117520,8),'logn')
grid
ylabel('EP quantiles')
xlabel('Lognormal quantiles')
title('(c)')
subplot(4,2,4)
gqqplot(EP_ready(100000:117520,8),'wbl')
grid
ylabel('EP quantiles')
xlabel('Weibull quantiles')
title('(d)')
subplot(4,2,5)
gqqplot(EP_ready(100000:117520,8),'gev')
grid
ylabel('EP quantiles')
xlabel('Generalized extreme value quantiles')
title('(e)')
subplot(4,2,6)
gqqplot(EP_ready(100000:117520,8),'gp')
grid
ylabel('EP quantiles')
xlabel('Generalized pareto quantiles')
title('(f)')
subplot(4,2,7)
gqqplot(EP_ready(100000:117520,8),'poiss')
grid
```

```
ylabel('EP quantiles')
xlabel('Poisson quantiles')
title('(g)')
subplot(4,2,8)
gqqplot(EP_ready(100000:117520,8),'rayl')
grid
ylabel('EP quantiles')
xlabel('Rayleigh quantiles')
title('(h)')
%-SP-----------------------------------------------
figure
subplot(4,2,1)
gqqplot(SP_ready(100000:117520,8),'norm')
grid
ylabel('SP quantiles')
xlabel('Normal quantiles')
title('(a)')
subplot(4,2,2)
gqqplot(SP_ready(100000:117520,8),'gam')
grid
ylabel('SP quantiles')
xlabel('Gamma quantiles')
title('(b)')
subplot(4,2,3)
gqqplot(SP_ready(100000:117520,8),'logn')
grid
ylabel('SP quantiles')
xlabel('Lognormal quantiles')
title('(c)')
subplot(4,2,4)
gqqplot(SP_ready(100000:117520,8),'wbl')
grid
ylabel('SP quantiles')
xlabel('Weibull quantiles')
title('(d)')
subplot(4,2,5)
gqqplot(SP_ready(100000:117520,8),'gev')
grid
ylabel('SP quantiles')
xlabel('Generalized extreme value quantiles')
title('(e)')
subplot(4,2,6)
gqqplot(SP_ready(100000:117520,8),'gp')
grid
ylabel('SP quantiles')
xlabel('Generalized pareto quantiles')
title('(f)')
subplot(4,2,7)
```

```
gqqplot(SP_ready(100000:117520,8),'poiss')
grid
ylabel('SP quantiles')
xlabel('Poisson quantiles')
title('(g)')
subplot(4,2,8)
gqqplot(SP_ready(100000:117520,8),'rayl')
grid
ylabel('SP quantiles')
xlabel('Rayleigh quantiles')
title('(h)')
%-CBD---------------------------------------------------
figure
subplot(4,2,1)
gqqplot(CBD_ready(100000:117520,8),'norm')
grid
ylabel('CBD quantiles')
xlabel('Normal quantiles')
title('(a)')
subplot(4,2,2)
gqqplot(CBD_ready(100000:117520,8),'gam')
grid
ylabel('CBD quantiles')
xlabel('Gamma quantiles')
title('(b)')
subplot(4,2,3)
gqqplot(CBD_ready(100000:117520,8),'logn')
grid
ylabel('CBD quantiles')
xlabel('Lognormal quantiles')
title('(c)')
subplot(4,2,4)
gqqplot(CBD_ready(100000:117520,8),'wbl')
grid
ylabel('CBD quantiles')
xlabel('Weibull quantiles')
title('(d)')
subplot(4,2,5)
gqqplot(CBD_ready(100000:117520,8),'gev')
grid
ylabel('CBD quantiles')
xlabel('Generalized extreme value quantiles')
title('(e)')
subplot(4,2,6)
gqqplot(CBD_ready(100000:117520,8),'gp')
grid
ylabel('CBD quantiles')
xlabel('Generalized pareto quantiles')
```

```
title('(f)')
subplot(4,2,7)
gqqplot(CBD_ready(100000:117520,8),'poiss')
grid
ylabel('CBD quantiles')
xlabel('Poisson quantiles')
title('(g)')
subplot(4,2,8)
gqqplot(CBD_ready(100000:117520,8),'rayl')
grid
ylabel('CBD quantiles')
xlabel('Rayleigh quantiles')
title('(h)')
%-CN-----------------------------------------------------
figure
subplot(4,2,1)
gqqplot(CN_ready(100000:117520,8),'norm')
grid
ylabel('CN quantiles')
xlabel('Normal quantiles')
title('(a)')
subplot(4,2,2)
gqqplot(CN_ready(100000:117520,8),'gam')
grid
ylabel('CN quantiles')
xlabel('Gamma quantiles')
title('(b)')
subplot(4,2,3)
gqqplot(CN_ready(100000:117520,8),'logn')
grid
ylabel('CN quantiles')
xlabel('Lognormal quantiles')
title('(c)')
subplot(4,2,4)
gqqplot(CN_ready(100000:117520,8),'wbl')
grid
ylabel('CN quantiles')
xlabel('Weibull quantiles')
title('(d)')
subplot(4,2,5)
gqqplot(CN_ready(100000:117520,8),'gev')
grid
ylabel('CN quantiles')
xlabel('Generalized extreme value quantiles')
title('(e)')
subplot(4,2,6)
gqqplot(CN_ready(100000:117520,8),'gp')
grid
```

```
ylabel('CN quantiles')
xlabel('Generalized pareto quantiles')
title('(f)')
subplot(4,2,7)
gqqplot(CN_ready(100000:117520,8),'poiss')
grid
ylabel('CN quantiles')
xlabel('Poisson quantiles')
title('(g)')
subplot(4,2,8)
gqqplot(CN_ready(100000:117520,8),'rayl')
grid
ylabel('CN quantiles')
xlabel('Rayleigh quantiles')
title('(h)')
%-CE-----------------------------------------------------
figure
subplot(4,2,1)
gqqplot(CE_ready(100000:117520,8),'norm')
grid
ylabel('CE quantiles')
xlabel('Normal quantiles')
title('(a)')
subplot(4,2,2)
gqqplot(CE_ready(100000:117520,8),'gam')
grid
ylabel('CE quantiles')
xlabel('Gamma quantiles')
title('(b)')
subplot(4,2,3)
gqqplot(CE_ready(100000:117520,8),'logn')
grid
ylabel('CE quantiles')
xlabel('Lognormal quantiles')
title('(c)')
subplot(4,2,4)
gqqplot(CE_ready(100000:117520,8),'wbl')
grid
ylabel('CE quantiles')
xlabel('Weibull quantiles')
title('(d)')
subplot(4,2,5)
gqqplot(CE_ready(100000:117520,8),'gev')
grid
ylabel('CE quantiles')
xlabel('Generalized extreme value quantiles')
title('(e)')
subplot(4,2,6)
```

```
gqqplot(CE_ready(100000:117520,8),'gp')
grid
ylabel('CE quantiles')
xlabel('Generalized pareto quantiles')
title('(f)')
subplot(4,2,7)
gqqplot(CE_ready(100000:117520,8),'poiss')
grid
ylabel('CE quantiles')
xlabel('Poisson quantiles')
title('(g)')
subplot(4,2,8)
gqqplot(CE_ready(100000:117520,8),'rayl')
grid
ylabel('CE quantiles')
xlabel('Rayleigh quantiles')
title('(h)')
%-CS-------------------------------------------------
figure
subplot(4,2,1)
gqqplot(CS_ready(100000:117520,8),'norm')
grid
ylabel('CS quantiles')
xlabel('Normal quantiles')
title('(a)')
subplot(4,2,2)
gqqplot(CS_ready(100000:117520,8),'gam')
grid
ylabel('CS quantiles')
xlabel('Gamma quantiles')
title('(b)')
subplot(4,2,3)
gqqplot(CS_ready(100000:117520,8),'logn')
grid
ylabel('CS quantiles')
xlabel('Lognormal quantiles')
title('(c)')
subplot(4,2,4)
gqqplot(CS_ready(100000:117520,8),'wbl')
grid
ylabel('CS quantiles')
xlabel('Weibull quantiles')
title('(d)')
subplot(4,2,5)
gqqplot(CS_ready(100000:117520,8),'gev')
grid
ylabel('CS quantiles')
xlabel('Generalized extreme value quantiles')
```

```
title('(e)')
subplot(4,2,6)
gqqplot(CS_ready(100000:117520,8),'gp')
grid
ylabel('CS quantiles')
xlabel('Generalized pareto quantiles')
title('(f)')
subplot(4,2,7)
gqqplot(CS_ready(100000:117520,8),'poiss')
grid
ylabel('CS quantiles')
xlabel('Poisson quantiles')
title('(g)')
subplot(4,2,8)
gqqplot(CS_ready(100000:117520,8),'rayl')
grid
ylabel('CS quantiles')
xlabel('Rayleigh quantiles')
title('(h)')
%-CG-----------------------------------------------------
figure
subplot(4,2,1)
gqqplot(CG_ready(100000:117520,8),'norm')
grid
ylabel('CG quantiles')
xlabel('Normal quantiles')
title('(a)')
subplot(4,2,2)
gqqplot(CG_ready(100000:117520,8),'gam')
grid
ylabel('CG quantiles')
xlabel('Gamma quantiles')
title('(b)')
subplot(4,2,3)
gqqplot(CG_ready(100000:117520,8),'logn')
grid
ylabel('CG quantiles')
xlabel('Lognormal quantiles')
title('(c)')
subplot(4,2,4)
gqqplot(CG_ready(100000:117520,8),'wbl')
grid
ylabel('CG quantiles')
xlabel('Weibull quantiles')
title('(d)')
subplot(4,2,5)
gqqplot(CG_ready(100000:117520,8),'gev')
grid
```

```
ylabel('CG quantiles')
xlabel('Generalized extreme value quantiles')
title('(e)')
subplot(4,2,6)
gqqplot(CG_ready(100000:117520,8),'gp')
grid
ylabel('CG quantiles')
xlabel('Generalized pareto quantiles')
title('(f)')
subplot(4,2,7)
gqqplot(CG_ready(100000:117520,8),'poiss')
grid
ylabel('CG quantiles')
xlabel('Poisson quantiles')
title('(g)')
subplot(4,2,8)
gqqplot(CG_ready(100000:117520,8),'rayl')
grid
ylabel('CG quantiles')
xlabel('Rayleigh quantiles')
title('(h)')
```

vi. DATA PREPARATION FOR TRAINING

```
%-------------------------------------------------------------
-------------
%Title: Data preparation for training puposes
%-------------------------------------------------------------
-------------
%Date: 07-02-2012
%-------------------------------------------------------------
-------------
%Status: fully functional
%-------------------------------------------------------------
-------------
%Notes: the input of this code can be any matrix with the
label of _ready.
%The full list of probable input variables are NP_ready (from
MN file),
%EP_ready (from ME file), SP_ready (from MS file), CBD_ready
(from CBD
%file), CN_ready (from CN file), CE_ready (from CE file),
CS_ready (from CS
%file), , CG_ready (from CG file)
%-------------------------------------------------------------
-------------
clear all;
```

```
clc;
close all;
% the user should change this part of the code based on the
file that
% he/she want to load
%attention: the user must change the last 3 lines accordingly
load('MN');
a=NP_ready; %a must have 8 columns at this stage and the
following is the explanation of each
% 1 year
% 2 month
% 3 day of the month
% 4 hour oh the day in 24 hours format
% 5 weekday, in a way that sun is 1, mon is 2,.... and sat is
7
% 6 temperature in degrees celsius
% 7 relative humidity
% 8 load in MW
temp1=a(:,3);
%we want the following 12 inputs this time
% the target variable is load
targets=a(:,8);
% X1 year
X1=a(:,1);
% X2 Month
X2=a(:,2);
% X3 hour
X3=a(:,4);
% X4 Weekday
X4=a(:,5);
% X5 temp
X5=a(:,6);
% X6 Humidity
X6=a(:,7);
% X7 previous week same hour load
X7= [NaN(168*2,1); targets(1:end-168*2)];
% X8 previous day same hour load
X8= [NaN(24*2,1); targets(1:end-24*2)];
% X9 avg month_day load at that half an hour ---useless---
worse results
%instead of that we are going to use this input
% X9 previous half an hour load
X9=[NaN(1,1);targets(1:end-1)];
% X10 is holiday
%-----holidays---X10
%sat and sun
X10=zeros(length(a),1);
for counter=1:length(a)
```

```
  if X4(counter)==1
    X10(counter)=1;
  end
  if X4(counter)==7
    X10(counter)=1;
  end
end
%-------wa public holidays importing
[num, text] = xlsread('waholidays.xls');
holidays = text(1:end,1);
holidaycounter=5; %the first holiday after our data starts

for counter=1:length(a)
  [x1 x2 x3]= datevec(holidays(holidaycounter),'mm/dd/yyyy');
  if (X1(counter)==x1) && (X2(counter)==x2) &&
(temp1(counter)==x3)
    midcounter=counter;
    check=0;
    while check==0
       if midcounter>=length(a)
         check=1;
       end
       X10(midcounter)=1;
       if X3(midcounter)==23.5;
         check=1;
       end
       midcounter=midcounter+1;
    end
    holidaycounter=holidaycounter+1;
  end
end

% X11 Average load in the previous 24 hours
X11=filter(ones(1,48)/48, 1, targets); %Average load of
previous 24 hours X11(n)=mean(targets(n-47:n))


% X12 Average load in the previous seven days
X12=filter(ones(1,336)/336, 1, targets); %Average load of
previous week X12(n)=mean(targets(n-335:n))

NP_X=[X1,X2,X3,X4,X5,X6,X7,X8,X9,X10,X11,X12];
NP_targets=targets;
save('MN','NP_raw','NP_double','NP_without_outlier','NP_ready'
,'NP_X','NP_targets')
```

vii. Neural Network and Decision Tree Set Up

```
%-------------------------------------------------------------
-------------
%Title: Nueral network and decision tree set up
%-------------------------------------------------------------
-------------
%Date: 11-08-2011
%-------------------------------------------------------------
-------------
%Status: has been tested and is working properly
%-------------------------------------------------------------
------------
clear all;
clc;
close all;
% the user should change this part of the code based on the
file that
% he/she want to load
%attention: the user must change the last line accordingly
%1
%load('MN');
%2
%load('ME');
%3
%load('MS');
%4
%load('CBD')
%5
%load('CN');
%6
%load('CE');
%7
load('CS');
%8
%load('CG');
X1=X(:,1);
X2=X(:,2);
X3=X(:,3);
X4=X(:,4);
X5=X(:,5);
X6=X(:,6);
X7=X(:,7);
X8=X(:,8);
X9=X(:,9);
X10=X(:,10);
X11=X(:,11);
X12=X(:,12);
temp1=X(:,13);
```

```
%selecting one year as test year
inputTRAIN=[X1(1:end-17520) X2(1:end-17520) X3(1:end-17520)
X4(1:end-17520) X5(1:end-17520) X6(1:end-17520) X7(1:end-
17520) X8(1:end-17520) X10(1:end-17520) X11(1:end-17520)
X12(1:end-17520)];
inputTEST=[X1(length(inputTRAIN)+1:end)
X2(length(inputTRAIN)+1:end) X3(length(inputTRAIN)+1:end)
X4(length(inputTRAIN)+1:end) X5(length(inputTRAIN)+1:end)
X6(length(inputTRAIN)+1:end) X7(length(inputTRAIN)+1:end)
X8(length(inputTRAIN)+1:end) X10(length(inputTRAIN)+1:end)
X11(length(inputTRAIN)+1:end) X12(length(inputTRAIN)+1:end)];
temp1TEST=temp1(length(inputTRAIN)+1:end);

targetTRAIN=targets(1:end-17520);
targetTEST=targets(length(inputTRAIN)+1:end);

%-------------------------------------------------------------
-----------
%Nueral Network Training
net = newfit(inputTRAIN', targetTRAIN', 40);
net.performFcn = 'mae';
net = train(net, inputTRAIN', targetTRAIN');

forecastLoadNN = sim(net, inputTEST')';
%-------------------------------------------------------------
---------
%bootstrap aggregated regression trees training

model=TreeBagger(40,inputTRAIN,targetTRAIN,'method','regressio
n','minleaf',30);
model=compact(model)

%Simulating the forecasts for both methods

forecastLoadNN = sim(net, inputTEST')';
forecastLoadRT = predict(model, inputTEST);

%1
%save('MN','NP_raw','NP_double','NP_without_outlier','NP_ready
','X','targets','net','model','forecastLoadNN','forecastLoadRT
')
%2
%save('ME','EP_raw','EP_double','EP_without_outlier','EP_ready
','X','targets','net','model','forecastLoadNN','forecastLoadRT
')
%3
```

```
%save('MS','SP_raw','SP_double','SP_without_outlier','SP_ready
','X','targets','net','model','forecastLoadNN','forecastLoadRT
')
%4
%save('CBD','CBD_raw','CBD_double','CBD_without_outlier','CBD_
ready','X','targets','net','model','forecastLoadNN','forecastL
oadRT')
%5
%save('CN','CN_raw','CN_double','CN_without_outlier','CN_ready
','X','targets','net','model','forecastLoadNN','forecastLoadRT
')
%6
%save('CE','CE_raw','CE_double','CE_without_outlier','CE_ready
','X','targets','net','model','forecastLoadNN','forecastLoadRT
')
%7
save('CS','CS_raw','CS_double','CS_without_outlier','CS_ready'
,'X','targets','net','model','forecastLoadNN','forecastLoadRT'
)
%8
%save('CG','CG_raw','CG_double','CG_without_outlier','CG_ready
','X','targets','net','model','forecastLoadNN','forecastLoadRT
')
```

viii. MAPE CALCULATOR

```
%-------------------------------------------------------------
------------
%title: MAPE table generator
%-------------------------------------------------------------
------------
%date: 09-02-2012
%modified on: 13-02-2012
%-------------------------------------------------------------
------------
%status: has been tested and is working properly
%-------------------------------------------------------------
----------

clc
% to calculate MAPE
disp('-----------------------------');
disp('NN MAPE');
disp('-----------------------------');
%Jan
msg = sprintf('MAPE of Jan: %f ',mean(abs(targetTEST(1:48*31-
1)-forecastLoadNN(1:48*31-1)))/mean(targetTEST(1:48*31-1)));
```

```
disp(msg);
%Feb
msg = sprintf('MAPE of Feb: %f',mean(abs(targetTEST(48*31-
1:48*(31+28)-1)-forecastLoadNN(48*31-1:48*(31+28)-
1)))/mean(targetTEST(48*31-1:48*(31+28)-1)));
disp(msg);
%March
msg = sprintf('MAPE of Mar:
%f',mean(abs(targetTEST(48*(31+28)-1:48*(2*31+28)-1)-
forecastLoadNN(48*(31+28)-1:48*(2*31+28)-
1)))/mean(targetTEST(48*(31+28)-1:48*(2*31+28)-1)));
disp(msg);
%April
msg = sprintf('MAPE of Apr:
%f',mean(abs(targetTEST(48*(2*31+28)-1:48*(2*31+28+30)-1)-
forecastLoadNN(48*(2*31+28)-1:48*(2*31+28+30)-
1)))/mean(targetTEST(48*(2*31+28)-1:48*(2*31+28+30)-1)));
disp(msg);
%May
msg = sprintf('MAPE of May:
%f',mean(abs(targetTEST(48*(2*31+28+30)-1:48*(3*31+28+30)-1)-
forecastLoadNN(48*(2*31+28+30)-1:48*(3*31+28+30)-
1)))/mean(targetTEST(48*(2*31+28+30)-1:48*(3*31+28+30)-1)));
disp(msg);
%June
msg = sprintf('MAPE of Jun:
%f',mean(abs(targetTEST(48*(3*31+28+30)-1:48*(3*31+28+2*30)-
1)-forecastLoadNN(48*(3*31+28+30)-1:48*(3*31+28+2*30)-
1)))/mean(targetTEST(48*(3*31+28+30)-1:48*(3*31+28+2*30)-1)));
disp(msg);
%July
msg = sprintf('MAPE of Jul:
%f',mean(abs(targetTEST(48*(3*31+28+2*30)-1:48*(4*31+28+2*30)-
1)-forecastLoadNN(48*(3*31+28+2*30)-1:48*(4*31+28+2*30)-
1)))/mean(targetTEST(48*(3*31+28+2*30)-1:48*(4*31+28+2*30)-
1)));
disp(msg);
%Aug
msg = sprintf('MAPE of Aug:
%f',mean(abs(targetTEST(48*(4*31+28+2*30)-1:48*(5*31+28+2*30)-
1)-forecastLoadNN(48*(4*31+28+2*30)-1:48*(5*31+28+2*30)-
1)))/mean(targetTEST(48*(4*31+28+2*30)-1:48*(5*31+28+2*30)-
1)));
disp(msg);
%Sep
msg = sprintf('MAPE of Sep:
%f',mean(abs(targetTEST(48*(5*31+28+2*30)-1:48*(5*31+28+3*30)-
1)-forecastLoadNN(48*(5*31+28+2*30)-1:48*(5*31+28+3*30)-
```

```
1)))/mean(targetTEST(48*(5*31+28+2*30)-1:48*(5*31+28+3*30)-
1)));
disp(msg);
%Oct
msg = sprintf('MAPE of Oct:
%f',mean(abs(targetTEST(48*(5*31+28+3*30)-1:48*(6*31+28+3*30)-
1)-forecastLoadNN(48*(5*31+28+3*30)-1:48*(6*31+28+3*30)-
1)))/mean(targetTEST(48*(5*31+28+3*30)-1:48*(6*31+28+3*30)-
1)));
disp(msg);
%Nov
msg = sprintf('MAPE of Nov:
%f',mean(abs(targetTEST(48*(6*31+28+3*30)-1:48*(6*31+28+4*30)-
1)-forecastLoadNN(48*(6*31+28+3*30)-1:48*(6*31+28+4*30)-
1)))/mean(targetTEST(48*(6*31+28+3*30)-1:48*(6*31+28+4*30)-
1)));
disp(msg);
%Dec
msg = sprintf('MAPE of Dec:
%f',mean(abs(targetTEST(48*(6*31+28+4*30)-1:48*(7*31+28+4*30)-
1)-forecastLoadNN(48*(6*31+28+4*30)-1:48*(7*31+28+4*30)-
1)))/mean(targetTEST(48*(6*31+28+4*30)-1:48*(7*31+28+4*30)-
1)));
disp(msg);
%-------------------------------------
disp('-----------------------------');
disp('DT MAPE');
disp('-----------------------------');
%Jan
msg = sprintf('MAPE of Jan: %f ',mean(abs(targetTEST(1:48*31-
1)-forecastLoadRT(1:48*31-1)))/mean(targetTEST(1:48*31-1)));
disp(msg);
%Feb
msg = sprintf('MAPE of Feb: %f',mean(abs(targetTEST(48*31-
1:48*(31+28)-1)-forecastLoadRT(48*31-1:48*(31+28)-
1)))/mean(targetTEST(48*31-1:48*(31+28)-1)));
disp(msg);
%March
msg = sprintf('MAPE of Mar:
%f',mean(abs(targetTEST(48*(31+28)-1:48*(2*31+28)-1)-
forecastLoadRT(48*(31+28)-1:48*(2*31+28)-
1)))/mean(targetTEST(48*(31+28)-1:48*(2*31+28)-1)));
disp(msg);
%April
msg = sprintf('MAPE of Apr:
%f',mean(abs(targetTEST(48*(2*31+28)-1:48*(2*31+28+30)-1)-
forecastLoadRT(48*(2*31+28)-1:48*(2*31+28+30)-
1)))/mean(targetTEST(48*(2*31+28)-1:48*(2*31+28+30)-1)));
```

```
disp(msg);
%May
msg = sprintf('MAPE of May:
%f',mean(abs(targetTEST(48*(2*31+28+30)-1:48*(3*31+28+30)-1)-
forecastLoadRT(48*(2*31+28+30)-1:48*(3*31+28+30)-
1)))/mean(targetTEST(48*(2*31+28+30)-1:48*(3*31+28+30)-1)));
disp(msg);
%June
msg = sprintf('MAPE of Jun:
%f',mean(abs(targetTEST(48*(3*31+28+30)-1:48*(3*31+28+2*30)-
1)-forecastLoadRT(48*(3*31+28+30)-1:48*(3*31+28+2*30)-
1)))/mean(targetTEST(48*(3*31+28+30)-1:48*(3*31+28+2*30)-1)));
disp(msg);
%July
msg = sprintf('MAPE of Jul:
%f',mean(abs(targetTEST(48*(3*31+28+2*30)-1:48*(4*31+28+2*30)-
1)-forecastLoadRT(48*(3*31+28+2*30)-1:48*(4*31+28+2*30)-
1)))/mean(targetTEST(48*(3*31+28+2*30)-1:48*(4*31+28+2*30)-
1)));
disp(msg);
%Aug
msg = sprintf('MAPE of Aug:
%f',mean(abs(targetTEST(48*(4*31+28+2*30)-1:48*(5*31+28+2*30)-
1)-forecastLoadRT(48*(4*31+28+2*30)-1:48*(5*31+28+2*30)-
1)))/mean(targetTEST(48*(4*31+28+2*30)-1:48*(5*31+28+2*30)-
1)));
disp(msg);
%Sep
msg = sprintf('MAPE of Sep:
%f',mean(abs(targetTEST(48*(5*31+28+2*30)-1:48*(5*31+28+3*30)-
1)-forecastLoadRT(48*(5*31+28+2*30)-1:48*(5*31+28+3*30)-
1)))/mean(targetTEST(48*(5*31+28+2*30)-1:48*(5*31+28+3*30)-
1)));
disp(msg);
%Oct
msg = sprintf('MAPE of Oct:
%f',mean(abs(targetTEST(48*(5*31+28+3*30)-1:48*(6*31+28+3*30)-
1)-forecastLoadRT(48*(5*31+28+3*30)-1:48*(6*31+28+3*30)-
1)))/mean(targetTEST(48*(5*31+28+3*30)-1:48*(6*31+28+3*30)-
1)));
disp(msg);
%Nov
msg = sprintf('MAPE of Nov:
%f',mean(abs(targetTEST(48*(6*31+28+3*30)-1:48*(6*31+28+4*30)-
1)-forecastLoadRT(48*(6*31+28+3*30)-1:48*(6*31+28+4*30)-
1)))/mean(targetTEST(48*(6*31+28+3*30)-1:48*(6*31+28+4*30)-
1)));
disp(msg);
```

```
%Dec
msg = sprintf('MAPE of Dec:
%f',mean(abs(targetTEST(48*(6*31+28+4*30)-1:48*(7*31+28+4*30)-
1)-forecastLoadRT(48*(6*31+28+4*30)-1:48*(7*31+28+4*30)-
1)))/mean(targetTEST(48*(6*31+28+4*30)-1:48*(7*31+28+4*30)-
1)));
disp(msg);
%-------------------------------------
disp('-----------------------------');
disp('Average Temperture');
disp('-----------------------------');
%Jan
msg = sprintf('Average Temperature of Jan: %.1f
',nanmean(X5(1:48*31-1)));
disp(msg);
%Feb
msg = sprintf('Average Temperature of Feb: %.1f
',nanmean(X5(48*31-1:48*(31+28)-1)));
disp(msg);
%Mar
msg = sprintf('Average Temperature of Mar: %.1f
',nanmean(X5(48*(31+28)-1:48*(2*31+28)-1)));
disp(msg);
%Apr
msg = sprintf('Average Temperature of Apr: %.1f
',nanmean(X5(48*(2*31+28)-1:48*(2*31+28+30)-1)));
disp(msg);
%May
msg = sprintf('Average Temperature of May: %.1f
',nanmean(X5(48*(2*31+28+30)-1:48*(3*31+28+30)-1)));
disp(msg);
%Jun
msg = sprintf('Average Temperature of Jun: %.1f
',nanmean(X5(48*(3*31+28+30)-1:48*(3*31+28+2*30)-1)));
disp(msg);
%Jul
msg = sprintf('Average Temperature of Jul: %.1f
',nanmean(X5(48*(3*31+28+2*30)-1:48*(4*31+28+2*30)-1)));
disp(msg);
%Aug
msg = sprintf('Average Temperature of Aug: %.1f
',nanmean(X5(48*(4*31+28+2*30)-1:48*(5*31+28+2*30)-1)));
disp(msg);
%Sep
msg = sprintf('Average Temperature of Sep: %.1f
',nanmean(X5(48*(5*31+28+2*30)-1:48*(5*31+28+3*30)-1)));
disp(msg);
%Oct
```

```
msg = sprintf('Average Temperature of Oct: %.1f
',nanmean(X5(48*(5*31+28+3*30)-1:48*(6*31+28+3*30)-1)));
disp(msg);
%Nov
msg = sprintf('Average Temperature of Nov: %.1f
',nanmean(X5(48*(6*31+28+3*30)-1:48*(6*31+28+4*30)-1)));
disp(msg);
%Dec
msg = sprintf('Average Temperature of Dec: %.1f
',nanmean(X5(48*(6*31+28+4*30)-1:48*(7*31+28+4*30)-1)));
disp(msg);
```

ix. SWIS MAP GENERATOR

```
function [ ] = SWISLoadMap(
CountryNorthLoad,CountryEastLoad,CountrySouthLoad,CountryGoldfieldsLoad,Met
roLoad ,DateandTime)
%SWISLoadMap gets the load in the Swiss-all region and generate the
%  Date Created: 22-09-2011
%proper map presentation of the load data
% the input variables are the load of different regions in the following
% order
% Country North
% Country East
% Country South
% Country Goldfields
% Metro
%-------------------------------------------------------
%METROLoadMap is a simillar function for metro area
%-------------------------------------------------------
load SWIS_ALL
figure1=figure;
patch(oneX,oneY,0,'EdgeColor',[0 0 .5625]);
patch(twoX,twoY,0,'EdgeColor',[0 0 .5625]);
patch(indianoceanX,indianoceanY,0,'EdgeColor',[0 0 .5625]);
a=patch(countrygoldfieldsX,countrygoldfieldsY,CountryGoldfieldsLoad,'EdgeCo
lor',[0 0 .5625]);
b=patch(countryeastX,countryeastY,CountryEastLoad,'EdgeColor',[0 0 .5625]);
c=patch(countrynorthX,countrynorthY,CountryNorthLoad,'EdgeColor',[0 0
.5625]);
d=patch(countrysouthX,countrysouthY,CountrySouthLoad,'EdgeColor',[0 0
.5625]);
e=patch(metroX,metroY,MetroLoad,'EdgeColor',[0 0 .5625]);
colorbar
legend([a,b,c,d,e],'Country Goldfields','Country East','Country
North','Country South','Metro');
title({'Spatial load forecast of SWIS region on' DateandTime});
axis tight

% Create textbox
annotation(figure1,'textbox',...
  [0.331802221886336 0.41969696969697 0.0442558192261187
0.0891837449640916],...
```

```
  'String',{MetroLoad 'MW'},...
  'FitBoxToText','off');

% Create textbox
annotation(figure1,'textbox',...
  [0.3213125 0.727272727272727 0.0728833887545345 0.032842879663688],...
  'String',{CountryNorthLoad 'MW'},...
  'FitBoxToText','off');

% Create textbox
annotation(figure1,'textbox',...
  [0.5181875 0.466763005780347 0.0839890417170496 0.0447976878612716],...
  'String',{CountryEastLoad 'MW'},...
  'FitBoxToText','off');

% Create textbox
annotation(figure1,'textbox',...
  [0.40725 0.238439306358381 0.0865
0.0491329479768784],'String',{CountrySouthLoad 'MW'},...
  'FitBoxToText','off');

% Create textbox
annotation(figure1,'textbox',...
  [0.690864003929868 0.501515151515151 0.0467418001813777
0.0682940970397602],...
  'String',{CountryGoldfieldsLoad 'MW'},...
  'FitBoxToText','off');

end
```

x. METROPOLITAN MAP GENERATOR

```
function [ ] = METROLoadMap(
MetroNorthLoad,MetroEastLoad,MetroSouthLoad,MetroCBDLoad,DateandTime )
%SWISLoadMap gets the load in the Swiss-all region and generate the
%  Date Created: 22-09-2011
%proper map presentation of the load data
% the input variables are the load of different regions in the following
% order
% Metro North
% Metro East
% Metro South
% Metro CBD
%-------------------------------------------------------
%SWISLoadMap is a simillar function for the whole SWIS area
%-------------------------------------------------------
load SWIS_METRO
%x=x-50;
%y=440-(y-440);
figure1=figure;
patch([0 526 526 0], [0 0 880 880],0,'EdgeColor',[0 0 .5625])
% patch(threeX,threeY,0,'EdgeColor',[0 0 .5625]);
% patch(fourX,fourY,0,'EdgeColor',[0 0 .5625]);
b=patch(metroeastX,metroeastY,MetroEastLoad,'EdgeColor',[0 0 .5625]);
```

```
c=patch(metronorthX,metronorthY,MetroNorthLoad,'EdgeColor',[0 0 .5625]);
d=patch(metrosouthX,metrosouthY,MetroSouthLoad,'EdgeColor',[0 0 .5625]);
e=patch(cbdX,cbdY,MetroCBDLoad,'EdgeColor',[0 0 .5625]);
colorbar
legend([b,c,d,e],'Metro East','Metro North','Metro South','CBD');
title({'Spatial load forecast of Perth metropolitan region on'
DateandTime});
axis tight

% Create textarrow
annotation(figure1,'textarrow',[0.423956931359354 0.429340511440108],...
  [0.589551181102362 0.508661417322835],'TextEdgeColor','none',...
  'String',{MetroCBDLoad 'MW'});

% Create textbox
annotation(figure1,'textbox',...
  [0.389963660834455 0.588976377952756 0.0905208613728129
0.0283464566929138],...
  'FitBoxToText','off');

% Create textbox
annotation(figure1,'textbox',...
  [0.359008075370121 0.74488188976378 0.0945585464333782
0.0330708661417311],...
  'String',{MetroNorthLoad 'MW'},...
  'FitBoxToText','off');

% Create textbox
annotation(figure1,'textbox',...
  [0.546433378196501 0.498212598425197 0.0969044414535666
0.0440944881889764],...
  'String',{MetroEastLoad 'MW'});

% Create textbox
annotation(figure1,'textbox',...
  [0.434724091520861 0.337582677165354 0.0820995962314939
0.0440944881889764],...
  'String',{MetroSouthLoad 'MW'});
End
```

Appendix

E. NOMENCLATURE

C_{FOSM}	Fixed plant operation and maintenance costs in $/MW.year
C_{POSM}	Variable plant operation and maintenance costs in $/MW.hour
C_{VOSM}	Capital investment cost of generation unit in $
C_{inv}	Plant fuel cost in $/MW
C_{FOSM}	Maximum allowable number of generation units of type j
C_{POSM}	X coordinate of the nearest load centre to generation unit i
C_{VOSM}	X coordinate of generation unit i
C_{inv}	Y coordinate of the nearest load centre to generation unit i
C_{fuel}	Y coordinate of generation unit i
num_j	Number of generation units of type j
x_{coni}	Boolean regional exclusive variable
x_{unit_i}	
y_{coni}	
y_{unit_i}	
num_j	
w_{xy}	
h	Hour counter
B	Maximum capital investment considered for all the new generation units
G_{max}	Maximum unit generation capacity MW/hour
GWP	Global warming potential
D_{max}	Maximum predicted demand of a load node MW/hour
N_{max}	Maximum number of types of generation units
N_{max}	Maximum number of generation units
N_{max}	Objective function
N_{con}	Number of consumption nodes
N_{opt}	Optimisation horizon in years
N_{opt}	GHG emission in tons/MW
N_{opt}	Generation output of a unit in MW/hour
h	Hour counter
N_{gen}	Generation unit node counter
N_{gen}	Type of generation unit
N_{gen}	Consumption node counter
N_{gen}	Actual electrical load in MW
N_{gen}	Forecasted electrical load in MW
N_{gen}	Spinning reserve (a positive number less than 1)
N_{gen}	Year counter
N_{gen}	Generation plant annual full load equivalent operational hours

Appendix

F. GLOSSARY OF-TERMS

ABARE	Australian Bureau of Agricultural and Resource Economics
ABM	Agent based model
ABS	Australian Bureau of Statistics
AC	Alternating current
Aerosols	Fine liquid or solid particles suspending in a fluid. In this document aerosols are dealing with solid or liquid particles that are floating in the atmosphere like smoke, air pollution, smog, and tear gas.
Anthropogenic activities	Activities caused by humans like burning fossil fuels and deforestation
ARX	Auto-regression with exogenous variables
BOM	Bureau of Meteorology
CCGT	Combined cycle generation technology
CHP	Combined heat and power plant
DC	Direct current
Deregulated electricity market	A deregulated market is owned and controlled by different individuals and enterprises. In this kind of market participants have more freedom to decide and plan for their facilities.
DG	Distributed generation
Distribution substation	A distribution substation transfers power from the transmission system to the distribution system of an area. Transformers inside distribution substation reduce the voltage level from transmission level to distribution level. They are also used to isolate distribution lines for maintenance and fault clearance purposes.
Elitism	Elitism is copying the best solutions to the next generation. This procedure helps the solution not to lose the best individuals.
G&T	Generation and transmission
GA	Genetic algorithm
GDP	Gross domestic product
GEV	Generalised extreme value distribution
GHG	Green house gas
Global warming potential	The estimate of global warming contribution of a given mass of a greenhouse gas. GWP is a relative measure of different GHGs in comparison with CO_2. GWP of CO_2 is 1.
GP	Generalised Pareto distribution

Green house gases	The atmosphere gases that absorb and emit radiation within the thermal infrared range. Most dominant atmospheric green house gases are water vapour and carbon dioxide.
Gross domestic product	Gross domestic product is a factor of country overall economical output. GDP is closely related with the standard of living. It is the market value of all final goods and services made within a country in one year.
GWP	Global warming potential
HVDC	High voltage direct current
IPCC	Intergovernmental Panel on Climate Change
LMP	Locational marginal pricing
Locational marginal pricing	The Locational Marginal Price (LMP) is a market-pricing approach used to manage the efficient use of the transmission system when congestion occurs on the bulk power grid.
Locus	The position of a gene on the chromosome
LRET	Large scale renewable energy target
LTLF	Long-term load forecast / forecasting
MAPE	Mean absolute percentage error
Meshing	Meshing is the act of dividing the area under study into small sections. Detailed study of different regions, their interactions and their effect on the whole area characteristics can be investigated in a more efficient manner by meshing.
Monopolistic electricity market	Monopoly in electricity market exists when a specific individual or enterprise controls the market and makes decisions for all components.
MTLF	Medium-term load forecast / forecasting
NIEIR	National institute of economy and industrial research
Nitrogen oxide	Nitrogen oxide can refer to these compounds of nitrogen and oxygen: nitric oxide (NO), nitrogen dioxide (NO_2), nitrous oxide (N_2O), nitrate radical (NO_3), nitrogen(VI) oxide, dinitrogen trioxide (N_2O_3), nitrogen(II,IV) oxide, dinitrogen tetroxide (N_2O_4), and dinitrogen pentoxide (N_2O_5).
Pareto-optimal	A set of optimum solutions
Pearson correlation coefficient	Pearson correlation coefficient shows the linear relationship between two variables. This coefficient varies from -1 to +1. A correlation of +1 shows the perfect positive linear relationships between two variables. Similarly, -1 correlation means perfect negative linear relationship and 0 means no relationship between two variables.
PV	Photovoltaic
RE	Renewable energy

RET	Renewable energy target
Search Space	All possible solutions to the problem
Spatial load forecasting	Future prediction of electricity demand geographical distribution
Spinning reserve	A generating capacity that can be injected to the network in few seconds in case of any failure occurrence at another generation unit or in case of sudden increment in load. It is impossible for a generator to reach its full power and synchronised with grid in few seconds and they should be in spinning mode.
SRET	Small scale renewable energy target
Statistical regression	The relation between selected values of x and observed values of y (that helps to predict the future value of y in a given value of x).
STLF	Short-term load forecast / forecasting
SWIS	South West Interconnected System (interconnected electricity grid of the South West of Western Australia)
T&D	Transmission and distribution
Trait	Possible aspect of an individual
Transmission congestion	Transmission congestion occurs when there is insufficient energy to meet the demands of all customers.
Transmission substation	A transmission substation connects two or more transmission lines. Some of them may have transformers to change voltage level but they are mainly used to isolate transmission lines for maintenance and fault clearance purposes.
UNEP	United Nations Environment Program
WMO	World Metrological Organisation